

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 14:59:26 ; Search time 34.67 seconds
(without alignments)
1681.965 Million cell updates/sec

Title: DAVISN-P04196.PPP

Perfect score: 3036

Sequence: 1 MKALIAALLITLQYSCAVS.....GKFKSGPQVSMFTTFPPK 525

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	323.5	10.7	245	22	ABG28019 Novel human diagno
2	318.5	10.5	330	22	ABG25331 Novel human diagno
3	311	10.2	491	22	ABG09949 Novel human diagno
4	306	10.1	351	13	RAR24393 Sequence of Histid
5	296.5	9.8	406	22	ABG27250 Novel human diagno
6	284.5	9.4	233	22	ABG19767 Novel human diagno
7	283.5	9.3	292	22	ABG32108 Peptide #4759 enco
8	283.5	9.3	292	22	ABG37360 Peptide #4866 enco
9	283.5	9.3	292	22	ABG22646 Protein #4645 enco
10	283.5	9.3	292	22	AAW58050 Human brain expres
11	283.5	9.3	292	22	AAW70490 Human bone marrow

12	283.5	9.3	292	22	AAW18326 Peptide #4760 enco
13	283.5	9.3	292	22	AAW30814 Peptide #4851 enco
14	283.5	9.3	292	22	AAW05933 Peptide #4615 enco
15	283	9.3	183	22	ABG25337 Novel human diagno
16	273.5	9.0	598	22	ABG23595 Novel human diagno
17	263	8.7	324	22	ABG20509 Novel human diagno
18	256	8.4	594	22	ABG1362 Drosophila melanog
19	251.5	8.3	227	22	ABG14399 Novel human diagno
20	250.5	8.3	304	22	ABG08412 Novel human diagno
21	234	7.7	124	22	ABG12879 Novel human diagno
22	226.5	7.5	388	22	ABW51345 Mouse HS-glycoprot
23	224.5	7.4	1401	22	ABG24037 Novel human diagno
24	224	7.4	382	20	AAW88491 Human liver clone
25	224	7.4	382	21	AAW25782 Human secreted pro
26	224	7.4	382	22	AAW51346 Human HS-glycoprot
27	224	7.4	382	22	AAW75368 Human secreted pro
28	222.5	7.3	372	22	ABG67014 Drosophila melanog
29	217.5	7.2	167	22	ABG64598 Drosophila melanog
30	217	7.1	266	22	ABG03494 Human peptide #14
31	217	7.1	485	22	ABW59608 Peptide #15 encode
32	217	7.1	644	22	ABG21101 Protein #14 encode
33	213.5	7.0	102	22	ABG68218 Human brain expres
34	211	6.9	449	22	ABW71938 Human bone marrow
35	210	6.9	82	22	AAW13580 Peptide #14 encode
36	210	6.9	82	22	ABW27363 Peptide #15 encode
37	210	6.9	82	22	ABW2509 Protein #14 encode
38	210	6.9	82	22	AAW53343 Human brain expres
39	210	6.9	82	22	AAW65721 Human bone marrow
40	210	6.9	82	22	AAW13580 Peptide #14 encode
41	210	6.9	82	22	AAW25978 Peptide #15 encode
42	210	6.9	82	22	AAW01333 Peptide #15 encode
43	207.5	6.8	79	22	AAU32905 Novel human secret
44	204.5	6.7	378	22	AAW51347 Bovine HS-glycopro
45	202.5	6.7	940	22	AAU28194 Novel human secret

ALIGNMENTS

RESULT 1
ABG28019
ID ABG28019 standard; Protein; 245 AA.
XX AC
XX ABG28019;
XX AC
DT 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #28010.
XX Human; chromosome mapping; gene mapping; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS92206.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess

Query Match 10.1%; Score 306; DB 13; Length 351;
Best Local Similarity 34.0%; Pred. No. 4.4e-19;
Matches 66; Conservative 10; Mismatches 86; Indels 32; Gaps 8;

QY 246 PQEHENINGVPHLGHPPFWGCHGERSSTTKPFKPHGSRDHHH-PHKPHGHPPPPPDER 304
Db 98 phphhhhhhhphh--hnhhghhhhh-----haahhhhhhhhhhaahhhhhhee 147
QY 305 DSHSGPPLPQGPPLLPMSCSQCHATFTGNGAQRSHNNNSDLHPKHHSHEQHPHGH 364
Db 148 hhhhhhaahhp-----wfhhhghyhhhaaphhhhhhhha---phhhhhhhaphh 197
QY 365 H-----PHAHHPHEH--DTHRHQPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 417
Db 198 hnhhhaphhhhhhhhaphhhhhhhhghhhhhhhhhghhhhhhhhhghhhhhhhhhhd---- 253
QY 418 DPPPHNOGHCCHGH 431
Db 254 ahhhhhhhhdahhh 267

RESULT 5
ID ABG27250 standard; Protein; 406 AA.
XX AC ABG27250;
-XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27241.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS91437.
XX PS New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 57609; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
mapping and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 406 AA;

Query Match 9.8%; Score 296.5; DB 22; Length 406;
Best Local Similarity 33.7%; Pred. No. 3.8e-18;
Matches 59; Conservative 5; Mismatches 64; Indels 47; Gaps 4;

QY 276 PPFKPHGSRDHHHPKHHPHGHGPPPPPPDERDHSHPPLPQGP----- 317
Db 139 pppsp-----sppppppppppppppppppppppppppppppppppsl 191
QY 318 -PLPMSCSQCHATFTGNGAQRSHNNNSDLHPKHHSHEQHPHGHHPHHPHEHDT 376
Db 192 spslpppsp-----spbhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 242
QY 377 HRQHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 431
Db 243 hhh 285

RESULT 6
ID ABG19767 standard; Protein; 233 AA.
XX AC ABG19767;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #19758.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS83954.
XX PS New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 50126; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
mapping and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations

PN WO200157277-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00669.
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 PS Claim 27; SEQ ID NO 29995; 639pp + sequence listing; English.
 XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO.int/pub/published_pct_sequences.
 XX Sequence 292 AA;
 SQ
 Query Match 9.3%; Score 283.5; DB 22; Length 292;
 Best Local Similarity 26.5%; Pred. No. 3.6e-17;
 Matches 66; Conservative 9; Mismatches 85; Indels 89; Gaps 7;
 QY 261 HPFHGGERSTTKPPFKPGSRDHHHPKPEHG----- 296
 DB 53 hphhnhhhhhhhhyhyhyhhhhhhhhhhhhhhhhhhhhhhrrhypttttttip 112
 QY 297 -----pppppDERHS 307
 DB 113 titittptttstittittattittittittittittittittittittittttttttttip 170
 QY 308 HGPPLPQGPPLLPMSCCSQHATFTNGAQRSHNNSSDLHPKHHSHEQHPGHHPH 367
 DB 171 lsspsqspsp--plchchhyhly--lyyhhhhrrhhhyhhhhhhhhhhhyhhhh 226
 QY 368 AHHPEHDTTHROPHGHCHHHPGHCHHP-----HGHHPHCHHCHDFODYGCDDP 420
 DB 227 hhhphhhhhhhhyhlyh--hyhphhyhpnlyhyhchhqqh-hlphhhhyhfhlyh 283
 QY 421 PHNOGHCH 429
 DB 284 hhhhlhhsch 292
 RESULT 9
 ID ABB22646
 XX ABB22646 standard; Protein; 292 AA.
 AC ABB22646;
 DT 23-JAN-2002 (first entry)
 XX Protein #4645 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 XX congenital heart disease.
 OS Homo sapiens.
 XX WO200157274-A2.
 PN 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00666.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488999/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 PS Claim 15; SEQ ID NO 24416; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 292 AA;
 SQ
 Query Match 9.3%; Score 283.5; DB 22; Length 292;
 Best Local Similarity 26.5%; Pred. No. 3.6e-17;
 Matches 66; Conservative 9; Mismatches 85; Indels 89; Gaps 7;
 QY 261 HPFHGGERSTTKPPFKPGSRDHHHPKPEHG----- 296
 DB 53 hphhnhhhhhhhhyhyhyhhhhhhhhhhhhhhhhhhhhhhrrhypttttttip 112
 QY 297 -----pppppDERHS 307
 DB 113 titittptttstittittattittittittittittittittittittttttttttttip 170
 QY 308 HGPPLPQGPPLLPMSCCSQHATFTNGAQRSHNNSSDLHPKHHSHEQHPGHHPH 367
 DB 171 lsspsqspsp--plchchhyhly--lyyhhhhrrhhhyhhhhhhhhhhhyhhhh 226
 QY 368 AHHPEHDTTHROPHGHCHHHPGHCHHP-----HGHHPHCHHCHDFODYGCDDP 420
 DB 227 hhhphhhhhhhhyhlyh--hyhphhyhpnlyhyhchhqqh-hlphhhhyhfhlyh 283
 QY 421 PHNOGHCH 429
 DB 284 hhhhlhhsch 292
 RESULT 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:01:21 ; Search time 15.34 Seconds
(without alignments)
835,948 Million cell updates/sec

Title: DAVISM-P04196.PEP
Perfect score: 3036
Sequence: 1 MKALIAALLITLQYCAVS.....GPKSGPPQVSMFFTHFPK 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217.5	7.2	167	4	Sequence 13, Appl
2	198.5	6.5	796	4	Sequence 40, Appl
3	187.5	6.2	858	4	Sequence 41, Appl
4	187.5	6.2	892	4	Sequence 42, Appl
5	185	6.1	26	3	Sequence 1, Appl
6	185	6.1	26	3	Sequence 2, Appl
7	169	5.6	1185	4	Sequence 23, Appl
8	165.5	5.5	309	1	Sequence 2, Appl
9	160.5	5.3	1958	1	Sequence 2, Appl
10	157.5	5.2	189	1	Sequence 6, Appl
11	153	5.0	89	4	Sequence 12, Appl
12	147.5	4.9	306	1	Sequence 12, Appl
13	144.5	4.8	60	1	Sequence 1, Appl
14	144.5	4.8	60	2	Sequence 1, Appl
15	144.5	4.8	60	5	Sequence 1, Appl
16	141.5	4.7	285	1	Sequence 7, Appl
17	137.5	4.5	285	1	Sequence 5, Appl
18	134.5	4.4	313	3	Sequence 3, Appl
19	134.5	4.4	313	4	Sequence 3, Appl
20	134.5	4.4	337	3	Sequence 2, Appl
21	134.5	4.4	337	4	Sequence 2, Appl
22	134	4.4	398	4	Sequence 17, Appl
23	129.5	4.3	402	4	Sequence 5, Appl
24	128.5	4.2	474	4	Sequence 10, Appl
25	126.5	4.2	285	1	Sequence 10, Appl
26	126	4.2	552	4	Sequence 7, Appl
27	126	4.2	960	4	Sequence 6, Appl

28	125.5	4.1	255	4	US-09-613-126-1	Sequence 1, Appl
29	125.5	4.1	265	1	US-08-292-045-2	Sequence 2, Appl
30	125.5	4.1	509	4	US-08-860-635A-21	Sequence 21, Appl
31	125.5	4.1	509	4	US-09-281-476-21	Sequence 21, Appl
32	125	4.1	349	4	US-09-461-474-12	Sequence 12, Appl
33	125	4.1	960	4	US-09-219-849-5	Sequence 5, Appl
34	124.5	4.1	826	1	US-07-638-431-2	Sequence 2, Appl
35	124.5	4.1	826	5	PCT-US92-00018-2	Sequence 2, Appl
36	124	4.1	1182	4	US-09-041-886-21	Sequence 21, Appl
37	123.5	4.1	448	4	US-09-461-474-8	Sequence 8, Appl
38	123.5	4.1	507	4	US-08-860-635A-19	Sequence 19, Appl
39	123.5	4.1	507	4	US-09-281-476-19	Sequence 19, Appl
40	123	4.1	367	2	US-08-737-045-14	Sequence 14, Appl
41	123	4.1	367	3	US-08-932-871B-2	Sequence 2, Appl
42	123	4.1	367	3	US-09-476-919-2	Sequence 2, Appl
43	123	4.1	367	4	US-08-780-311A-2	Sequence 2, Appl
44	122.5	4.0	124	2	US-08-925-237-2	Sequence 2, Appl
45	122.5	4.0	352	2	US-08-737-045-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-507-323B-13
; Sequence 13, Application US/09507323B
; Patent No. 6274345
; GENERAL INFORMATION:
; APPLICANT: Choi, Jong Hyun
; APPLICANT: Lee, Sang Yup
; APPLICANT: Xu, Zhaohui
; TITLE OF INVENTION: EXPRESSION VECTORS COMPRISING A GENE
; TITLE OF INVENTION: ENCODING OMPC DERIVED FROM E.COLI AS CELL SURFACE ANCHORING
; TITLE OF INVENTION: MOTIF
; FILE REFERENCE: HYLER39.001AUS
; CURRENT APPLICATION NUMBER: US/09/507,323B
; CURRENT FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: KR 10-1999-0005773
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Histidine linker.
US-09-507-323B-13

Query Match	7.2%	Score 217.5;	DB 4;	Length 167;
Best Local Similarity	30.7%	Pred. No. 3.5e-13;		
Matches	55;	Conservative	4;	Mismatches 67;
				Indels 53; Gaps 9;
QY	280	PHGSDRHHP-----HKPHEHGPPPPPPDERDHSHGPPPLPOGPPPLPMSCSCQAHATFG	333	
Db	5	PSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSG-----	46	
QY	334	TNGAQRHSNNNSDLHPEKHSH-----EQHPHGHH-----PHAHPHEHDTROHPHGH	384	
Db	47	-----HHHHHSGLDPSGHHHHHSGLDPSGHHHHHSGLDPSGHHHHS--GLDPSGH	99	
QY	385	HPHGH---HPHGHHPHGH---HPHGHHPHGHFDQYGCDDPPHNOGHCCHGHPGPH	437	
Db	100	HHHHHSGLDPSGHHHHHSGLDPSGHHHHH-----SGLDPSGHHHH--HSGLDPSGH	151	
RESULT 2				
US-08-857-076-40				
; Sequence 40, Application US/08857076C				
; Patent No. 6225120				
; GENERAL INFORMATION:				
; APPLICANT: Ruvkun, Gary				


```
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-08-857-076-42

Query Match      6.2%; Score 187.5; DB 4; Length 892;
Best Local Similarity 20.2%; Pred. No. 2.8e-09;
Matches 111; Conservative 63; Mismatches 182; Indels 193; Gaps 27;

QY 55 DAHLDRVENTVYLVLDVQESCSVL--SRKYNDCEPPDSRRPSEIVIGOCKVIATRH 112
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 154 DPDLFTKNTNTEYDLVDLVKLGKPAVDEARK--KIEVPDASAPKNKIVEYLMYYRTLK 210
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 113 SHESQDLRIDNCTTSSVSSALANTKDSPLIDFEDTERYRKQANKALEKYKEENDDF 172
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 211 ESE-----LIQLMAYRTKRNLSLNLVKNNI-----DREFDQACESLVKLLKDKKNDL 259
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 173 ASFRVDRIERVARVRGEGTGTF-----VDFSVRNCPRHFF----- 209
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 260 QNL-ID-----VLSKGTKTGCTITPTLDGRQVHGKRGFPVYVVGKLRNFNEMTKNET 314
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 210 RHPNVGFCRADLYDVEALDLESPK-----NLVINCEVDFDPOHENINGVPPHLGH 261
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 315 RHVD---HCK-----HAFEMKSDMVCVNPVHYEIVGTMIVGORDHNRDMPPH--Q 362
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 262 PFHWGHE-----RSSTTRPPKPH-----GSRDHHHPHK-PHEH 295
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 363 RYHTFGQDPVDMRSRIFPPAPPPMNMHTRPQPMQOLPSVGATFAHPLPHOAPHNP 422
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 296 GPPPPDERDHS-----GPPLOGPPPLPMSCCSQHATFTGTNGAQRHSHNNN 345
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 423 GVSHFYSIAPQTHYPLNMNPPIQMPQMPPPPL-----HOGYGMNGPSCSENHN 473
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 346 SSDLHPKHHSHEQHPHGHHPHHPHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGH 405
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 474 --PFQNHNYNDISHEN-HYSYDCGNLXGFTPTYPDFHHPFNQPH----- 517
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 406 PHCHDFQDYGCDPPPHNGCHCHGHPGPPHLLRRGPKGP--RPF----- 450
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 518 -----QPQLSQNHTSQGSHQPGH--QGQVPNDPPIRSRVLQPSTVTLDFVRR 564
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 451 HCRQI-----GSVYR-----LPLRKCEVLP--LPEAN 476
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 565 YCRQTFGNRFEGSESQGAIIIRSNKFTIEEFDSPICGVTVVYVRPMRTDGEVLNIMPE-- 622
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 477 FPSFPLPH 485
| : |
Db 623 ----DAPYH 627

RESULT 5
US-08-785-636-1
; Sequence 1, Application US/08785636
; Patent No. 6027942
; GENERAL INFORMATION:
```

```
; APPLICANT: Yip, Tai-Tung
; APPLICANT: Hutchens, T. William
; TITLE OF INVENTION: Method and Apparatus for Desorption and Ionization of
; TITLE OF INVENTION: Analyses
; FILE REFERENCE: Hutchens
; CURRENT APPLICATION NUMBER: US/08/785,636
; CURRENT FILING DATE: 1997-01-17
; EARLIER APPLICATION NUMBER: 08/068,896
; EARLIER FILING DATE: 1993-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: peptide
US-08-785-636-1

Query Match      6.1%; Score 185; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 GHHPGHHPGHHPGHHPGHHPGHHPH 407
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 GHHPGHHPGHHPGHHPGHHPGHHPH 25
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 6
US-09-095-407-2
; Sequence 2, Application US/09095407
; Patent No. 6124137
; GENERAL INFORMATION:
; APPLICANT: Hutchens, T. William
; TITLE OF INVENTION: Surface-Enhanced Neat Desorption for the
; TITLE OF INVENTION: Detection of Analytes
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fulbright & Jaworski L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77010-3095
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,407
; FILING DATE: 10-Jun-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/068,896
; FILING DATE: 28-MAY-1993
; APPLICATION NUMBER: WO PCT/US94/06064
; FILING DATE: 27-MAY-1994
; APPLICATION NUMBER: US 08/556,951
; FILING DATE: 27-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5639-PCT-US-D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 651-5325
; TELEFAX: (713) 651-5246
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: amino acid
```

```

; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-095-407-2

Query Match          6.1%; Score 185; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 GHHPHGHHPHGHHPHGHHPH 407
Db 1 GHHPHGHHPHGHHPHGHHPH 25

RESULT 7
US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredezen, Dale E.
; APPLICANT: Rabizadeh, Sharoz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-23

Query Match          5.6%; Score 169; DB 4; Length 1185;
Best Local Similarity 23.5%; Pred. No. 2.5e-07;
Matches 87; Conservative 27; Mismatches 127; Indels 130; Gaps 21;

QY 255 VPHLGHHPHWG-----GHER-----SSTTKPPFK-PHGRDRHH--- 288
Db 325 LPGLHSPYAMGQGMGLPPGPEKGTLPSPHSLPPASSAPAPMRFPYSSSSSSAA 384
QY 289 -----PHKPEHGGPPPPDERDHSHP-----LP-----OG 315
Db 385 ASSSSSSSSSAPFPASALPSYHSF---PPPTSLSVSNQPKYQTQPSLPSQAVWSQG 441
QY 316 PPLPLP-----MSCSCQHATF-GTNGAQRHSHNNSSDLPHKHHSHQHPHHPH 370
Db 442 PPPPPYGRLLANSNAHPGPPFPSTCAQSTAHPPVST--HHHHHQOQOQOQOQOQH 499

QY 371 PHEHDTHQHPHGHHPHGHHPH---GHHPHGHHPHGHHPHCHDFQDY--GPCD-PPPHNQ 424
Db 500 GNS-----GPPPPGAFPHPLEGSSHHAHYAMSPSLGSLRYPYPCGAHLPPHSQ 550
QY 425 GHCCHG--HGPPPGH-----LRRGPKGKGRPPFHCROIGSVYRLPLRK-- 466
Db 551 VSYSQAGNPGPPVSSSSSSSTSGSYPCSHFSPSQGQ-----GAPYPPFPVPTVT 603
QY 467 -----GEVLPPLP---EANFPSPFPLPHHKHPLKPDNQPPQSVSESCPKCKSGFP 513
Db 604 TSSATLSTVIATVASSPAGYKTASPPGPPYGK-----RAPSPGAYKTATPPGYKGPSP 657
QY 514 QVSMFFTHFTFP 524
Db 658 --PSFRTGTPP 666

RESULT 8
US-08-161-406-2
; Sequence 2, Application US/08161406
; Patent No. 5476785
; GENERAL INFORMATION:
; APPLICANT: Wellens, Thomas
; APPLICANT: Howard, Russell J.
; TITLE OF INVENTION: Recombinant DNA Clone Containng a
; TITLE OF INVENTION: Genomic Fragment of PfHRP-II Gene from Plasmodium
; TITLE OF INVENTION: Falciparum
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/161,406
; FILING DATE: 06-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1173-448P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-161-406-2

Query Match          5.5%; Score 165.5; DB 1; Length 309;
Best Local Similarity 27.3%; Pred. No. 7.9e-08;
Matches 50; Conservative 13; Mismatches 67; Indels 53; Gaps 14;

QY 268 HERSSTTKPPFKPHGRDHHHPKHPEHGGPPPPPPDERDHSHPPLPQGPPLPLMSCSSC 327
Db 25 HETQAVDDAHHAHVADAHHAHAHAH-----DAHHAH----- 59
QY 328 QHATFTNGAQRHSHNNSSDLPHKHHSHQHPHGH--PHAHHPHEHDTHQHPH---- 382
Db 60 -HA-----ADAH-HAHAHAADAH-HAHAHAADAH-HAHAHAADAH-HAADAHAHAHAD 108
```


QY 426 HCCHGH 431
|||||
Db 53 GCCHGH 58

Search completed: May 31, 2002, 15:03:36
Job time: 135 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:01:16 ; Search time 20.98 Seconds
(without alignments)
2404.523 Million cell updates/sec

Title: DAVISN-P04196.PEP

Perfect score: 3036

Sequence: 1 MKALIAALLLITQYSCAVS.....GKFKSGFPOVSMFFTHTPPK 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3036	100.0	525	1 KGHUGH	histidine-rich gly
2	1514.5	49.9	445	2 A60488	histidine-rich gly
3	306	10.1	351	1 KGZQHL	histidine-rich gly
4	269	8.9	735	2 T45059	hypothetical prote
5	266	8.8	140	2 A54523	histidine-rich pro
6	223.5	7.4	471	2 T33997	hypothetical prote
7	217	7.1	644	1 KGHUHI	kininogen, HMW pre
8	213	7.0	774	2 T39539	alpha-amylase homo
9	209.5	6.9	1002	2 T43236	carboxypeptidase C
10	202.5	6.7	1571	2 T14155	zinc finger protei
11	199.5	6.6	834	2 T29821	hypothetical prote
12	197	6.5	330	2 S22140	nodulin Eno2 - Se
13	197	6.5	620	2 S06733	hydroxyproline-ric
14	194.5	6.4	2715	2 T13049	eyelid - fruit fly
15	193.5	6.4	350	2 T55214	salivary proline-r
16	193	6.4	621	1 KGBOH1	kininogen, HMW I p
17	192.5	6.3	1560	2 T00080	hypothetical prote
18	192	6.3	1378	2 T30173	zinc finger protei
19	190	6.3	619	1 KGBOH2	kininogen, HMW II
20	187.5	6.2	940	2 J60291	FB19 protein - hum
21	186	6.1	309	2 S08343	nodulin precursor
22	186	6.1	436	2 T49714	MHC H-2K/t-w5-link
23	186	6.1	508	2 A45477	cAMP response elem
24	184.5	6.1	309	2 S10889	proline-rich prote
25	180	5.9	242	2 H82061	hypothetical prote
26	179.5	5.9	580	2 T10863	extensin precursor
27	179.5	5.9	657	2 A29454	knob-associated hi
28	178.5	5.9	639	2 A25486	kininogen, HMW I p
29	178	5.9	852	2 A34373	histidine-rich cal

ALIGNMENTS

RESULT 1

KGHUGH

histidine-rich glycoprotein precursor - human

N:Alternate names: HRG

C:Species: Homo sapiens (man)

C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 16-Jun-2000

C:Accession: A01287; S29669

R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.

Biochemistry 25, 2220-2225, 1986

A>Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nu

A:Reference number: A01287; MUID:86216149

A:Accession: A01287

A:Molecule type: mRNA

A:Residues: 1-525 <KOI>

A:Cross-references: GB:AB005803; MID:g2280513; PIDN:BAA21613.1; PID:g2280514

R:Hennis, B.; Havelaar, A.; Kluff, C.

submitted to the EMBL Data Library, October 1991

A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich gly

A:Reference number: S29669

A:Accession: S29669

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 214-247 <HEN>

A:Cross-references: EMBL:Z17218; MID:g32453; PIDN:CAAY8925.1; PID:g32454

C:Comment: Although its physiological function is not yet known, HRG does bind heme, din, and the lysine-binding site of plasminogen. On the basis of its homology with HM

lood coagulation cascade.

C:Comment: The amino half of this protein is homologous to the first two cystatin-lik

could not have inhibitory activity.

C:Comment: In addition to having a high histidine and proline content, this protein h

e-rich' region.

C:Genetics:

A:Gene: GDB:HRG

A:Cross-references: GDB:120055; OMIM:142640

A:Map position: 3q27-3q27

C:Superfamily: histidine-rich glycoprotein; cystatin homology

C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-525/Product: histidine-rich glycoprotein #status predicted <MAT>

F:19-131/Domain: cystatin homology <CY1>

F:140-246/Domain: cystatin homology <CY2>

F:276-321/Region: proline-rich

F:348-437/Region: histidine-rich

F:351-497/Region: proline-rich

F:63,125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:78-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 100.0%; Score 3036; DB 1; Length 525;

Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKALIAALLLITLQYSCAVPTDCSAVEPEAEKALDLINKRRRDGYLFOLLRIADAHLD 60
DB 1 MKALIAALLLITLQYSCAVPTDCSAVEPEAEKALDLINKRRRDGYLFOLLRIADAHLD 60
QY 61 VENTTVYLVLDVQSDCVLSRKYNWDCPPDSRRPEIVIGQCKVIATRHSHESQDL 120
DB 61 VENTTVYLVLDVQSDCVLSRKYNWDCPPDSRRPEIVIGQCKVIATRHSHESQDL 120
QY 121 VIDFNCTTSSVSALANTKDSPLVLDFFEDTERYRKQANKALEYKEENDDFASFRVDR 180
DB 121 VIDFNCTTSSVSALANTKDSPLVLDFFEDTERYRKQANKALEYKEENDDFASFRVDR 180
QY 181 ERVARVGGEGTYGVDFSVRNCPRHHPHNFVFCRADLFYDVEALDLESPKLNLYN 240
DB 181 ERVARVGGEGTYGVDFSVRNCPRHHPHNFVFCRADLFYDVEALDLESPKLNLYN 240
QY 241 CEVFDQOEHNINGVPPHGLHGFHGGHRSSTTKPPKPHGSRDHHHPKPHGPPPP 300
DB 241 CEVFDQOEHNINGVPPHGLHGFHGGHRSSTTKPPKPHGSRDHHHPKPHGPPPP 300
QY 301 PDERHSHGPPPLPQGPPLPMSCSQCATFTNGAQRHSHNNNSDLHPKHSHSQH 360
DB 301 PDERHSHGPPPLPQGPPLPMSCSQCATFTNGAQRHSHNNNSDLHPKHSHSQH 360
QY 361 PHGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 420
DB 361 PHGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 420
QY 421 PHNQHCCHGPPPPHLLRRRGGKPRFHCRCQIGSVYRLPLRKGEVLPPLPEANFSP 480
DB 421 PHNQHCCHGPPPPHLLRRRGGKPRFHCRCQIGSVYRLPLRKGEVLPPLPEANFSP 480
QY 481 PLPHHKHPLKPNQPPQSVSSCPGKFGPPQVSMFTTTPPK 525
DB 481 PLPHHKHPLKPNQPPQSVSSCPGKFGPPQVSMFTTTPPK 525
RESULT 2
A60488
histidine-rich glycoprotein - bovine (fragments)
N:Alternate names: autorosette inhibition factor
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1993 #sequence revision 23-Mar-1995 #text_change 07-Jul-1995
C:Accession: S35687; J2196; A60488
R:Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.
FEBS Lett. 328, 285-290, 1993
A:Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly
A:Reference number: S35687; MUID:93351678
A:Accession: S35687
A:Molecule type: protein
A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>
A:Note: 355-Gln and 368-Tyr were also found
R:Halkier, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.
Biochem. Biophys. Res. Commun. 200, 78-82, 1994
A:Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII
A:Reference number: J2196; MUID:94220160
A:Accession: J2196
A:Molecule type: protein
A:Residues: 1-23;35-54,'VK',57-1101,'R','TVGEYTEG',116,'N',118,'R',120-136;137-145;150-20
A:Experimental source: plasma
R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halkier, T.
Thromb. Res. 60, 385-396, 1990
A:Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.
A:Reference number: A60488; MUID:91196010
A:Accession: A60488
A:Molecule type: protein
A:Residues: 1-6,'X',8-15 <VES>
C:Comment: This protein is a single-chained plasma protein which participates in transgl
C:Superfamily: histidine-rich glycoprotein; cystatin homology
C:Keywords: glycoprotein; plasma
F:2-113/Domain: cystatin homology <CY1>
F:122-207/Domain: cystatin homology (fragments) <CY2>
F:7-424,60-71,87-108,165-346,180-203,258-288/Disulfide bonds: #status experimental

F:74,107,164,266/Binding site: carbohydrate (Asn) (covalent) #status experimental
Query Match 49.9%; Score 1514.5; DB 2; Length 445;
Best Local Similarity 57.0%; Pred. No. 6e-103;
Matches 293; Conservative 37; Mismatches 109; Indels 75; Gaps 9;
QY 18 AVSPDTCSAVEPEAEKALDLINKRRRDGYLFOLLRIADAHLDVENTTVYLVLDVQSD 77
DB 1 AVNPDTCDAVEPEAVRALDLINK-GRDGYLFOLLRIADAHLDKVESIAVYLVLDKVESD 59
QY 78 CSVLRSKYNWDCPPDSRRPEIVIGQCKVIATRHSHESQDLRVDFNCTTSSVSALAN 137
DB 60 CPVLSRKHWDCCLNVSRYPSETVIGQCKVIAITLAGDELVSVDNCTTSSVSALTN 119
QY 138 TKDSPLVLDFFEDTERYRKQANKALEYKEENDDFASFRVDRERVARVGGEGTYGV 197
DB 120 IIDSPLVLDFFEDTLY-----ENSDPAPFRM-----RARGEGTYSFLD 159
QY 198 FSVRNCPRHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 257
DB 160 FSVRNCSSHHFPRHSHIFGFCRADLFYDVEASDLTPKDIVTNCVEFN-LKRRRESVAOH 218
QY 258 HLGHFPHGGHRSSTTKPPKPHGSRDHHHPKPHGPPPPDERDHSHP-----P 311
DB 219 HLGRFPHSGEHSHPAGPPPKPSGSKDHGHPHSHYFRCPPPLEHKNHSDSPPPQARAP 278
QY 312 LPQGPPLPPLPMSCSQCATFTNGAQRHSHNNNSDLHPKHSHSQHHPHHPHHPH 371
DB 279 LPFPPLPGL-----RCPHPPFGTKGNHR-----PFDHSS 307
QY 372 HEHDTROHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 431
DB 308 DEH-----HNNHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 352
QY 432 GPPGHLRRRGGKPRFHCRCQIGSVYRLPLRKGEVLPPLPEANFSPFLPHKHPKLP 491
DB 353 -PPRHSKERGPGKGFHFRHWRPTGYIHLPSLKKAGEVLPPLPEANFSPFLPHNLP 411
QY 492 DNQPPQSVSSCPGKFGPPQVSMFTTTPPK 525
DB 412 ETQAPQASCPGTFNFKFLHISKFFAYTLPK 445
RESULT 3
KGZQHL
histidine-rich glycoprotein precursor - plasmodium lophurae
C:Species: Plasmodium lophurae
C:Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A22692
R:Ravetch, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.
Nature 312, 616-620, 1984
A:Title: Primary structure and genomic organization of the histidine-rich protein of
A:Reference number: A22692; MUID:85061618
A:Accession: A22692
A:Molecule type: DNA
A:Residues: 1-351 <RAV>
A:Cross-references: GB:X01469; NID:g9997; PIDN:CAA25698.1; PID:g9999
C:Comment: There are two copies of 16-residue repeats, two copies of 17-residue repea
C:Genetics:
A:Introns: 23/3
C:Superfamily: plasmodium histidine-rich protein
C:Keywords: glycoprotein; tandem repeat
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-47/Domain: propeptide #status predicted <PRO>
F:48-351/Product: histidine-rich glycoprotein #status predicted <MAT>
F:59-74,75-90/Region: 16-residue repeats
F:91-107,108-123/Region: 17-residue repeats
F:124-138,139-153/Region: 15-residue repeats
F:173-301,312-331/Region: 10-residue repeats
F:40/Binding site: carbohydrate (Asn) (covalent) #status predicted

Db 3 LIATSLVPDEHTPMSPVNTTTLKORSGIKMEIPPLYDPSQDDPEDG-----VNPV 57
Qy 55 DAHLDRVENTVYVLLVDQSDCSVL---SRKYWDCRPPDSRSEIVIGOCKVIATRH 112
Db 58 DPDLFDTKNTNTEVDLVLGRPAVDEAR---KIEVPASAPPKNKIVELYMYRTLK 114
Qy 113 SHESODLRVIDFNCTTSVSSALANTKDSPLVIDFEDETERYKQANKALEKYKEENDF 172
Db 115 ESE-----LIQLNAYRTKRNLNLVKNNT-----DREFDQACESLVKRLKDKKNDL 163
Qy 173 ASERVD-----RIERVARVRGEGTGTFV-----DFSVR 201
Db 164 QNL-IDVVLSTGTYGTGITTPTLDGLQVHGRKRPVHYVYGLWRNEMTKNETRHVD 222
Qy 202 NCPRHFRPHRNVFG-FCRADLFDVEALDLESPKNLVINEVEFDPPQSHENINGVPPHLG 260
Db 223 HC-KHAFEMKSDMGEVCNYPHYEI-----VIGTMVQORDHNDMPPH-- 268
Qy 261 HPFWHGHE-----RSSTTKPPFKPH-----GSRDHHHPKH-PHE 294
Db 269 QRYHTPGRQDPVDDMSREIPPASIRPPPMNHTRPQMPQOLPSVYGATFAHPLHQAPHN 328
Qy 295 HGPPPPDPRDHS-----GPPLPQGGPPLPMSCCSCOHATFTGNGAQRHSHNN 344
Db 329 PGVSHPSYIAQTHYPLNMNIPQMPQMPPL-----HOGYGMNGPSCSENN 379
Qy 345 NSSDLHPKHHSHEQHPRGHGHHPHAHHPHEHDTHQHPRGHGHHPHGHHPHGH 404
Db 380 N-PHQNHNYNDISHPN-HYSYDCGNLYGFTPTYPDFHHPNQPH----- 424
Qy 405 HPHCHDFQDYGCDPPPHNQHCCHGCGPPPGHLLRRRGPGKGP--RP----- 449
Db 425 -----QPPQLSQNHTSQGSHOPGH-QGVNPDPPIISRPAALRTILLFIHD 470
Qy 450 -FHCRQIGSVYRLPLRKGEVLPL 472
Db 471 TFFCLRLSSFLKFKVLEVNSITVL 494
RESULT 12
S22140
nodulin Enod2 - Sesbania rostrata
C:Species: Sesbania rostrata
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C:Accession: S22140
R:de Bruijn, F.J.
submitted to the EMBL Data Library, November 1991
A:Reference number: S22140
A:Accession: S22140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <DEB>
A:Cross-references: EMBL:X63339; NID:g21372; PIDN:CAA44939.1; PID:g21373
C:Superfamily: proline-rich protein 3
Query Match 6.5%; Score 197; DB 2; Length 330;
Best Local Similarity 26.5%; Pred. No. 5.4e-07;
Matches 71; Conservative 22; Mismatches 99; Indels 76; Gaps 14;
Qy 246 POEHENINGVPPHGLHPHFW-GGHERSSTTKPKPKHSGRHHHP-----HKPHEHGPP-- 298
Db 109 PPEYQ-----PPHENPPPEYQPPHE-----KPP--PEYQPPHEKPPPEYQPPHEKPPPEY 156
Qy 299 PPDPRDHSHPPLPQGGPPLPMSCCSCOHATFTGNGAQRHSHNNSSDLHPKHHSHE 358
Db 157 QPPHEKPPPEYQPPHEKPPPVPPPEYKPPH-----EKPPSYE 194
Qy 359 Q-HPHGHHHPHHPHEHDTHQHPRGHGHHPHGHGHHHPHGHGHHHPHGHGHHHPHGHGHH 417
Db 195 KPPPYEKPPHEKPPPEYKPPHEKPPPYDKPKPPHEKPPHEKPPHEKPPPEY--- 251
Qy 418 DPPPHNQHCCHGCGPPPGHLLRRRGPGKGRPFHCRQIGSVYRLPLRKGEVLPLPEANF 477

Db 252 -KPPHEK-----PPPEY---KPPHEKPP-----YEKPPHEK-----PPPEYK- 285
Qy 478 PSFPLPHHKHPLKPDNQPPQSVSESCEP 505
Db 286 -----PPHEKPPPEYPPPVKPPPEYK 308
RESULT 13
S06733
hydroxyproline-rich glycoprotein precursor - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C:Accession: S06733
R:Keller, B.; Lamb, C.J.
Genes Dev. 3, 1639-1646, 1989
A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein ge
A:Reference number: S06733; MUID:90128263
A:Accession: S06733
A:Molecule type: DNA
A:Residues: 1-620 <KEL>
A:Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein
Query Match 6.5%; Score 197; DB 2; Length 620;
Best Local Similarity 27.0%; Pred. No. 1.1e-06;
Matches 75; Conservative 16; Mismatches 99; Indels 88; Gaps 15;
Qy 273 TTKPFPKHGRDHHHPKHHPHGGPPPPPPDERDHSHPPLPQGGPPLPMSCCSCOHATF 332
Db 56 TTTTPSRGHVSPRIHAPPR---HAYPPP---SHGLPSPVSGGPP----- 94
Qy 333 GTNGAQRHSHNNSSDLHPKH---HHSHEQHPHGHGHHHPHHPHEHDTHQHP---HGH 384
Db 95 -----HRGHLPPSRGFNPPSPVISPSPSPYCAPPSHGHPGLPSHGQPPSPSHGH 148
Qy 385 HPHGHHHPHGH---PHGHHHPHGHHPHCHDFQDYGCDPPPHNQHCCHGCGPPPGHLLRRRG 442
Db 149 AP---PSGGHTPPRGQHPHPSHR-----RPSPPSR-----HGHPPPTTYAQ--- 185
Qy 443 PKGPRPHPCR-----QIGSVYRLPLRKGEVLPLP--EAFNPSFPL-----PHHKH---- 487
Db 186---PPPTIYSPQVQPPPTTSPPTTTHVQPTTSPSPSRGHQPPPTTHRHAPPTTHRHAPPT 243
Qy 488 -----PLKPDNQPPQSVSESCEPCKFKSGFPQVS 516
Db 244 HQPSPLRHLPPSPRRQPPPTTYSPPPPAYAQ---PQPS 279
RESULT 14
TI3049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: TI3049
R:Treisman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: TI3049
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2715 <TRE>
A:Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding

Query Match	6.43;	Score 194.5;	DB 2;	Length 2715;
Best Local Similarity	21.96;	Pred. No. le-05;		
Matches	84;	Conservative	30;	Mismatches 136;
				Indels 133;
				Gaps 18;
QY	245	DP	--QEHENI----	NGVPPHLGHFFHWGGHERSSTTKPPF-KPHGSRDHHHRRKPHHGCP 297
DB	54	DPS	IQQQQNVAPRHYGAPPPG-----	SGPGGPGDPAAVMYIHHHQQQQOHP 103
QY	298	PPPP	DHSHGPPLPQGGPPPLPMS	SSCOHATFTGNGAORSHNNSSDLH---PKKH 354
DB	104	PPPH	MQQQHHGGAP--PP-----	GGAEHAPGVKEEYTHLPPPHPH 145
QY	355	HSHEQH-----	PHGHHPAHHPHEHDTHQPHGHHPHGHHPHGH 399	
DB	146	PAYG	IADPNWDYRYGQPLPGGK	PQQQQPHPQQPPQPGGSGPNRPPQRYIPGQ 205
QY	400	HPGH	HHPHCHD-FDYGPCDPPHNOGHCH-----	GHGP 433
DB	206	PPQG	PTPLNSLLQSSNPPPPQHRYANTYDPQAAASAAAA	AAAAAQQAQGGPPPPGHGP 265
QY	434	PPGH	L-----RRRGKGRPPHCHQIG--SYVRLP	---LRKGEVLPPEA----- 475
DB	266	PPQH	QSPYGGQGGWAPPPRPY-SPQLG	SPQQYRTPTPTNTSRGQ-SPYPPAHGQNSG 323
QY	476	NFP	SFPL-----	PHHKPLKPDNQPPQSVSE 502
DB	324	SYSP	QQQQQQQQQQQQAQGGP	VPGGPPGTCGQPPQNTPTTSQSPYQRY-P 382
QY	503	SCPK	FKSGGFPQVSMFTHTTPK	525
DB	383	TPGL	PAGGSHRTAYSTHGYPE	405

RESULT 15
I55214
salivary proline-rich glycoprotein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Nov-1999
C:Accession: I55214
R:Miao, Y.
J. Biol. Chem. 228, 343-350, 1995
A:Title: cDNA cloning and characterization of rat salivary glycoproteins: Novel members
A:Reference number: I55214
A:Accession: I55214
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-350 <RES>
A:Cross-references: GB:L08134; NID:g986942; PIDN:AAA75405.1; PID:g986943
C:Genetics:
A:Gene: sgp158
C:Superfamily: proline-rich protein
C:Keywords: glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-350/Product: salivary proline-rich glycoprotein #status predicted <MAT>

Query Match	6.4%	Score 193.5;	DB 2;	Length 350;
Best Local Similarity	24.5%;	Pred. No. 1e-06;		
Matches	81; Conservative 27;	Mismatches 110;	Indels 113;	Gaps 19;

QY	255	VPPILGHFFHWGGHRSSTTKPKPKHGSRDHHHPKH - EHG-----PDPPPERDHS-	307
	:	:::	: :: : :
Db	30	LPH-----PDHGNOTQPRPHDIGNQ--OPRPHPDHGNOTQPRPHPDGNGQTQ	80
	:	:	:
QY	308	-----HG-----PPLP-----QGPPPLLPMSCSCQHATGTGAQRHSHNNSSD	348
	:	:	:
Db	81	PRRPHPDHGNQTQPRPHPDGNGQTQPRP-----HPDHGNTQTPRRPHPDGNGQ	130
	:	:	:
QY	349	LPHRK-----HHSHEHQPCNH-----PNAHHPHE-HDTHQRHGHGHHPHGHHPH	396
	:	: :: : :	: :: : :
Db	131	TQPRPPPGGNNHTPRRPHPDHGNOTQPRHPHDGNGQTQPRPHPDHGNGQTQPRRPH	190
	:	:	:
QY	397	-----HGHNHPGHNHPCHDFQDGFCDP-----PPH-NOGHCCHG	430

```

Db      191  DQGNQTPRRPHDHGNQTPRRPHDPDQGNQTPRRPHDPDQGNQTPRRPHDPDQGNQTP 250
Qy      431  HGPPPGHLRRRGCKGPRCHRCIQISVYRLPLRKGE-----VLP 471
Db      251  RPHPD---QGQNTQPRPH-PDLGNQTPRRPRFGQGNQTPRRPRFGQGNQTPRRP 305
Qy      472  LP-----EANFPSPPLHHKHLKPDNQPFQ 498
Db      306  LPGGGNQNHPSPLP--DHPPEDYQRTQ 334

```

Search completed: May 31, 2002, 15:03:14
Job time: 118 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:03:16 ; Search time 13.48 Seconds
(without alignments)
1507.994 Million cell updates/sec

Title: DAVISN-P04196.PEP

Perfect score: 3036

Sequence: 1 MKALTAALLITLQSCAVS.....GKFKSGFPQVSMFFTHTFPK 525

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3036	100.0	525	1	HRG_HUMAN
2	1725.5	56.8	526	1	HRG_RABIT
3	1327	43.7	396	1	HRG_BOVIN
4	306	10.1	351	1	HRPX_PLALO
5	235	7.7	378	1	FETB_RAT
6	226.5	7.5	388	1	FETB_MOUSE
7	226.5	7.5	661	1	KNG_MOUSE
8	221	7.3	382	1	FETB_HUMAN
9	217	7.1	644	1	KNG_HUMAN
10	211	6.9	449	1	CSUP_DROME
11	197	6.5	620	1	EXTN_TOBAC
12	193	6.4	621	1	KNH1_BOVIN
13	190	6.3	619	1	KNH2_BOVIN
14	186	6.1	309	1	NO75_SOYBN
15	186	6.1	476	1	KE4_MOUSE
16	179.5	5.9	657	1	KNOB_PLAFN
17	178.5	5.9	639	1	KNG_RAT
18	178	5.9	852	1	SRCH_RABIT
19	175.5	5.8	515	1	KE4L_CAEEL
20	174.5	5.7	251	1	PRP2_HUMAN
21	174	5.7	276	1	PRPL_HUMAN
22	174	5.7	434	1	NO75_LOPLU
23	173	5.7	549	1	DSX_DROME
24	171.5	5.6	247	1	PRP4_HUMAN
25	171	5.6	352	1	KE4_BRARE
26	170.5	5.6	331	1	PRP1_HUMAN
27	169	5.6	1185	1	DRPL_HUMAN
28	168.5	5.6	1183	1	DRPL_RAT
29	168	5.5	234	1	PRPM_HUMAN
30	168	5.5	395	1	SRP_MOUSE
31	166.5	5.5	424	1	S3B4_HUMAN
32	165.5	5.5	332	1	HRP1_PLAFA
33	163	5.4	428	1	FXB2_MOUSE

RESULT 1

ID	HRG_HUMAN	STANDARD	PRT	525 AA.
AC	P04196;			
DT	20-MAR-1987 (Rel. 04, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).			
GN	HRG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86216149; PubMed=3011081;			
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;			
RT	"Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA.";			
RL	Biochemistry 25:2220-2225(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;			
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 214-247 FROM N.A.			
RX	MEDLINE=94245171; PubMed=8188234;			
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,			
RA	Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;			
RT	"Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29.";			
RL	Genomics 19:195-197(1994).			
RN	[4]			
RP	SEQUENCE OF 19-27.			
RC	TISSUE=Plasma;			
RX	MEDLINE=93092937; PubMed=1459097;			
RA	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,			
RA	Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,			
RA	Hochstrasser D.F.;			
RT	"Plasma protein map: an update by microsequencing.";			
RL	Electrophoresis 13:707-714(1992).			
CC	-!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS			
CC	HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE			
CC	FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,			
CC	AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS			
CC	HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN			
CC	MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD			
CC	COAGULATION CASCADE.			
CC	-!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS			
CC	PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATITIONS OF A 5-			
CC	RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.			
CC	-!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; M13149; AAA52694.1; -;
DR EMBL; AB005803; BAA21613.1; -;
DR EMBL; 217218; CAA78925.1; -;
DR PIR; A01287; KGHUGH.
DR SWISS-2DPAGE; P04196; HUMAN.
DR MIM; 142640; -;
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR Glycoprotein; Heparin-binding; Repeat; Signal.
FT SIGNAL 1 18
FT CHAIN 19 525 HISTIDINE-RICH GLYCOPROTEIN.
FT DOMAIN 19 136 CYSTATIN-LIKE 1.
FT DOMAIN 137 254 CYSTATIN-LIKE 2.
FT DOMAIN 276 321 PRO-RICH.
FT DOMAIN 350 497 PRO/HIS-RICH.
FT DISULFID 24 504 BY SIMILARITY.
FT DISULFID 78 89 BY SIMILARITY.
FT DISULFID 105 126 BY SIMILARITY.
FT DISULFID 203 417 BY SIMILARITY.
FT DISULFID 218 241 BY SIMILARITY.
FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;

Query Match 100.0%; Score 3036; DB 1; Length 525;

Best Local Similarity 100.0%; Pred. No. 1.7e-195; Indels 0; Gaps 0;
Matches 525; Conservative 0; Mismatches 0;

QY 1 MKALIAALLIITLQYSCAVPTDCSAVEPEAEKALDLINKRRDGYLFOLLRIADAHLD 60
DB 1 MKALIAALLIITLQYSCAVPTDCSAVEPEAEKALDLINKRRDGYLFOLLRIADAHLD 60
QY 61 VENTTVYVLVDQESDCSVLSKYNWDCPPDSRPPSEIVIGQCKVIATRHSHESQDLR 120
DB 61 VENTTVYVLVDQESDCSVLSKYNWDCPPDSRPPSEIVIGQCKVIATRHSHESQDLR 120
QY 121 VIDFNCTTSVSSALANTKDSPLVIDFEDTERYRKQANKALEKYEKNDDEAFSRVDRI 180
DB 121 VIDFNCTTSVSSALANTKDSPLVIDFEDTERYRKQANKALEKYEKNDDEAFSRVDRI 180
QY 181 ERVARVGGEGTGYFVDFSVRNCPRHHPHHPHNVGFCRADLFYDVEALDLESPKMLVN 240
DB 181 ERVARVGGEGTGYFVDFSVRNCPRHHPHHPHNVGFCRADLFYDVEALDLESPKMLVN 240
QY 241 CEVFDPOEHENINGVPHLGHHPHGHGHSSTTKPPFKPHGSRDHHHPKHEHGP PPP 300
DB 241 CEVFDPOEHENINGVPHLGHHPHGHGHSSTTKPPFKPHGSRDHHHPKHEHGP PPP 300
QY 301 PDERDHSHPPLPOGPPPLPMSCCSQHATGTCNAQRSHNNSSDLHPKHSHSEHQH 360
DB 301 PDERDHSHPPLPOGPPPLPMSCCSQHATGTCNAQRSHNNSSDLHPKHSHSEHQH 360
QY 361 PHGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 420
DB 361 PHGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 420
QY 421 PHNQGHCHGHGPPPPGHLRRRPGKGRPFHCRQIGSVYRLPLRKEVLPPLPEANFPSP 480
DB 421 PHNQGHCHGHGPPPPGHLRRRPGKGRPFHCRQIGSVYRLPLRKEVLPPLPEANFPSP 480
QY 481 PLPHHKLPLKPDQPPQSVESPCGKFKSGFPQVSMFFTHTPPK 525

DB 481 PLPHHKLPLKPDQPPQSVESPCGKFKSGFPQVSMFFTHTPPK 525

RESULT 2

HRG_RABIT 2
ID HRG_RABIT STANDARD; PRT; 526 AA.
AC Q28640;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histidine-rich glycoprotein precursor (Histidine-proline rich
DE glycoprotein) (HPRG) (Fragment).
GN HRG.

OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.

RC TISSUE=Serum;

RX MEDLINE=9622917; PubMed=8639676;

RA Borza D.-B., Tatum F.M., Morgan W.T.;

RT "Domain structure and conformation of histidine-proline-rich
glycoprotein.";

RL Biochemistry 35:1925-1934(1996).

CC -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS
HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE
FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,
AND THE LYSINE-BINDING SITE OF PLASMINOGEN ON THE BASIS OF ITS
HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN
MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD
COAGULATION CASCADE.

CC -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
PROTEIN HAS MANY INTERNAL REPEATS. 15 TANDEM REPEATS OF A 5-
RESIDUE SEQUENCE (G[H/P][H/P]PH, CONSENSUS) FORM A HIS/PRO-RICH
REGION.

CC -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

CC EMBL; U32189; AAC48516.1; -;

CC InterPro; IPR000010; Cystatin.

CC Pfam; PF00031; cystatin; 1.

CC SMART; SM00043; CY; 2.

KW Glycoprotein; Heparin-binding; Repeat; Signal.

FT NON_TER 1 1

FT SIGNAL <1 8

FT CHAIN 9 526

FT DOMAIN 9 126

FT DOMAIN 127 243

FT DOMAIN 251 296

FT DOMAIN 329 498

FT DISULFID 14 505

FT DISULFID 68 79

FT DISULFID 95 116

FT DISULFID 193 415

FT DISULFID 207 230

FT DISULFID 272 302

FT CARBOHYD 115 115

FT CARBOHYD 192 192

FT CARBOHYD 240 240

FT CARBOHYD 310 310

FT CARBOHYD 485 485

FT SITE 303 304

FT SITE 421 422

FT SEQUENCE 526 AA; 58877 MW; 810F23D367D93D42 CRC64;

CC	HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN	
CC	MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD	
CC	COAGULATION CASCADE.	
-I	COAGULATION: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS	
CC	PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATS OF A 5-	
CC	RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.	
CC	-I SIMILARITY: CONTAINS 2 CYPSTATIN-LIKE DOMAINS.	
DR	PIR; S35687; S35687.	
DR	PIR; PF00031; cystatin; 1.	
KW	Glycoprotein; Heparin-binding; Repeat.	
FT	DOMAIN 1 102	CYSTATIN-LIKE 1.
FT	DOMAIN 103 169	CYSTATIN-LIKE 2.
FT	DOMAIN 191 238	PRO-RICH.
FT	DOMAIN 243 368	PRO/HIS-RICH.
FT	DISULFID 7 375	
FT	DISULFID 56 67	
FT	DISULFID 77 92	
FT	DISULFID 123 297	
FT	DISULFID 137 160	
FT	DISULFID 212 242	
FT	NON_CONS 52 53	
FT	CARBOHYD 70 70	N-LINKED (GLCNAC. . .).
FT	NON_CONS 71 72	
FT	NON_CONS 78 79	
FT	CARBOHYD 91 91	N-LINKED (GLCNAC. . .).
FT	NON_CONS 103 104	
FT	CARBOHYD 122 122	N-LINKED (GLCNAC. . .).
FT	NON_CONS 163 164	
FT	CARBOHYD 220 220	N-LINKED (GLCNAC. . .).
FT	NON_CONS 263 264	
FT	NON_CONS 303 304	
FT	VARIANT 86 86	S -> R.
FT	VARIANT 309 309	S -> Q.
FT	VARIANT 322 322	H -> Y.
FT	SEQUENCE 396 AA; 44470 MW; 128A822349D6E6FC CRC64;	
QY	Query Match	43.7%; Score 1327; DB 1; Length 396;
DB	Best Local Similarity	51.0%; Pred. No. 1.1e-81;
	Matches 262; Conservative 38; Mismatches 90; Indels 124; Gaps	
QY	18	AVPTDCSAYEPRAEKALDLINKRRRDGYLFOLLRTADAHLDRENTVYVLLDVQESD 77
DB	1	AVNPTGCDVPEVAVRALDLINK-GRDGYLFOLLRVADAHLDKVESTAVYLV-----ESD 55
QY	78	CSVLSRKYYWDCPPDSRRRSEIVIGOCKVIATRHSHESODLRVIDFNCTTSVSSALAN 137
DB	56	CPVLSRKHWDC-----LNVIVIGOCKLAG-----PEDLSVNDENCTTSVSSALTN 103
QY	138	TKDSPVLIDPFEDTERYRKQANKALEKYEENDDFASFRVDRIERVARRGSGTGYYFD 197
DB	104	-----MRARGGEGTSYFLD 117
QY	198	FSVRNCRPHHPHPNPNVGFGRADLFYDVDAALDLESPKLVINCEVDFPOEHENINGVPP 257
DB	118	FSVRNCSHHFPRH-HIFGFCRADLFYDVASDLTPKDIVTNCEVF-----HRRFSVQH 172
QY	258	HLGHPFFWGGHERSSTTKPPFKPHGSRDHHHPHKPHEHGPPPPPPDERDHSHP-----P 311
DB	173	HLGRPFHSGEHEHSPACRPPFKPSGKDCHGHPHESYNFRCPPLPKHNHSDSPFQARAP 232
QY	312	LPOGPPPLLPMSCCSOHATFGTNGAQRHSHNNSSDLHPKHHSHEQHPGHHPHAHP 371
DB	233	LPFPFPGGL-----RCPHPFPGTKG-----NHRP-----P 256
QY	372	HEHDTHQHPHGHHPHGHHPHGHHPHGHHPHCHDFQDYGCDPPPHNQHCCHG 431
DB	257	HDHSSD-----HHPHGHHPHGHHPHGHHPHGHHPDNDYDHGCDPPPHR----- 303
QY	432	GPPPHLRRGPGKGRPPFHCRTQIGSVYRPLPLRKGEVLPLPEANFSPFLPHHKPLK 491
DB	304	PPPHSRKRGPGKHGRFHWRTGYTHRLPSLKGVLPLPEANFSPFLPHNNPLQ 362

Query Match	56.8%;	Score 1725.5;	DB 1;	Length 526;
Best Local Similarity	60.6%;	Pred. No. 3.9e-108;		
Matches 328; Conservative	48;	Mismatches 122;	Indels	43;
Gaps				
QY	12	TLOYCAVSPTDCSAVEPEAEKALDLINKRRRDGYLFQLLRADAHLDVRVENTVYYVLV	71	
Db	2	TLOCWALTFTDCKTKPLAEKALDLINKWRDDGYLFQLLRVADAHLDGAESTVYYVLV	61	
QY	72	DVGESDCSVLSRKRYWNDDPPDSRRPSEIIVIGOCKVIATRHSHESODLVIDFNCTSSV	131	
Db	62	DVAKETDCSVLSRKHWDCCDPLTKRPSLDVIGOCKVIATRSDEYQTLRLDNFNCFTSSV	121	
QY	132	SSLALANTKDSPLVDIFEDPERYRKQANKALEXYKEENDDFASFVRDRTERVAVRGEGG	191	
Db	122	SSLALANTKDSPLVDFTEFDERPKSADKALEYKYSEAYASFRVDRVERTRVRKGGER	181	
QY	192	TGYFDVFSVRCNPRHHFPHPNPVFGCRADLFYDVEALDLESKNLVINCEVDFDPOEHEN	251	
Db	182	TNYVDFSVCNSRSHFHRRP-AFGECRADLSFDVEASNLENEDVIIISCVNFEEHGN	240	
QY	252	INGVPHLGHFPFWGGHERSSTTKPFKPHGSRDDHHHKPHKBHGPPPPPDERDHSHGPP	311	
Db	241	ISGFRPHLG-----KTPLGTGDSRDHHDHHPKHKFGCPPQEGEDFSEGPP	286	
QY	312	LPGGPPELLPMSCSSCOHATFTGTNGAQRSHNNNSSDL----HPKKHHSHHQPHGHHPH	367	
Db	287	LQGGTPLLSPFPFRCRHRFFGTNETRPHHHISVINIHRPPHGHHPGPPHGHHPH	346	
QY	368	A-----HH-----PHEHDTHQRPHGHHHPHGHHHPHGHHHPHGHHHC	408	
Db	347	GPPPHGHHPGPPPRHPHPHPGPPHGHHPHPGPPHGHHPHPGPPHGHHPHPG	406	
QY	409	HDFQDYGCPDPHPNOG---HCHGHGPPPGHLLRRGCGKGRPRPHCQIGSVYRLPPL	464	
Db	407	HGFHDHGCDPPSHKEGPQDLH-QHAMGPPPKHPGRGPKGHFPFHWRRIAGSVYOLPPL	465	
QY	465	RKGEVLPLEANFPSPFLPHHKHKLXPDNQPQQSVSESQCFKFKSGFPQVSMFFTHTFP	524	
Db	466	QKGEVLPLEANFPQLLRNRHTPLKPEIQPPFOVASERCPFEENGFAQLSKFFPSTFP	525	
QY	525	K 525		
Db	526	K 526		
RESULT	3			
HRG_BOVIN				
ID	HRG_BOVIN	STANDARD;	PRT;	396 AA.
AC	F33433;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Histidine-rich glycoprotein (Histidine-proline rich glycoprotein)			
DE	(hERG) (Fragments).			
GN	HRG.			
OC	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE=93351678; PubMed=8348977;			
RA	Scoerensen C.B., Krogh-pedersen H., Petersen T.E.;			
RT	"Determinat-ion of the disulphide bridge arrangement of bovine			
RT	histidine-rich glycoprotein."			
RL	FEBS Lett. 328:285-290(1993).			
CC	-!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS			
CC	HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE			
CC	FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,			
CC	AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS			


```
Db 120 KAM-----FHVKNRRLVLPAYNCTLRVSKRKHSMCPDPCPHVPL--SAPSVEAAT 173
QY 160 KALEYKKE--DDFASFRVRIARVARGEGTGYFVDFSVR-----NCPRHIF 208
Db 174 ESLAKFNENPSQALVAVTK-----ATTQWVGSYFVEYDIKESPTQSDSCSLOAS 229
QY 209 PRHPNVGFCRADLYDVEALDLESP-----KNLVINCEVDFDQHEHNINGVPPHILG 260
Db 230 DSEP--VGLCOGSL-----IKSPGVPPQRFKKTIVVSCFEFSQDQ-----VP---- 270
QY 261 HPFHGCHGERST---TKPPFK---PHGSRDHHHPKHEHGP--PPPPDERDHDHGHGPP 312
Db 271 -----GGENPADTQAKLPQKNTAPTSPSITAPRGSIQHLPEQEPEDSKGSPPEPF 325
QY 313 P 313
Db 326 P 326

RESULT 6
FETB_MOUSE
ID FETB_MOUSE STANDARD; PRT; 388 AA.
AC Q90XCL1
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Fetus-B precursor (IRL685).
GN FETUB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=Liver;
RX MEDLINE=20407138; PubMed=10947975;
RA Olivier E., Soury E., Ruminy P., Husson A., Parmentier F., Daveau M.,
RA Salier J.-P.;
RT "Fetus-B, a second member of the fetus family in mammals.";
RL Biochem. J. 350:589-597(2000).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- TISSUE SPECIFICITY: LIVER, LUNG AND TONGUE.
CC -!- SIMILARITY: BELONGS TO THE FETUIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DB EMBL; AJ242927; CAB62541.1; -
DR MGD; MGI:1890221; Fetus.
DR InterPro; IPR00010; Cystatin.
DR InterPro; IPR001363; Fetusin.
DR Pfam; PF00031; cystatin; 2.
DR SMART; SM00043; Cy; 2.
DR PROSITE; PS01254; FETUIN_1; 1.
DR PROSITE; PS01255; FETUIN_2; 1.
KW Glycoprotein; signal; Repeat.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 388 FETUIN-B.
FT DOMAIN 27 152 CYSTATIN-LIKE 1.
FT DOMAIN 153 276 CYSTATIN-LIKE 2.
FT DISULFID 96 107 BY SIMILARITY.
FT DISULFID 120 140 BY SIMILARITY.
FT DISULFID 154 157 BY SIMILARITY.
FT DISULFID 217 224 BY SIMILARITY.
FT DISULFID 237 263 BY SIMILARITY.
FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).
```

```
FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 388 AA; 42712 MW; 39CFAD76A9D8DC2B CRC64;

Query Match 7.5%; Score 226.5; DB 1; Length 388;
Best Local Similarity 26.1%; Pred. No. 2.4e-08;
Matches 91; Conservative 52; Mismatches 135; Indels 71; Gaps 17;

QY 4 LIAALLLITLQYSC-----AVSPDTC--SAVEPEAEKALDLINKRRRDG 45
Db 3 LRLRLVLTACACMARSPAPLPQRPPLSLPLHPLGCCNDSEVLAVAGFALQINRDKDG 62
QY 46 YLFQLLRADHLDRVNT--TVYLLVDYQSDCSVLSKRYWNCDEPPSRPSEIVGQ 104
Db 63 YMLSLNRVHDVREHYQEDMGSLFYLTLDVLETDCHVLSRKAQDKCP--RMFYESVYQG 119
QY 105 KVIATRHSHESODLRVI---DFNCTTSSVSALANT---KDSPLVDLDFDTERYRKQAN 159
Db 120 KRAM-----FHINKRRVLYLPAYNCTLRVSKRKHHTTCTDCPSPIDLSPNSA--LEAAT 173
QY 160 KALEYKKEENDDFASFRVRIARVARGEGTGYFVDFSVR-----NCPRHIFPR 210
Db 174 ESLAKFNENPSQALVAVTK-----ATTQWVGSYFVEYDIKESPTQSDSCSLOHSDS 231
QY 211 HPNVGFCRADLYDVEALDLESP---KNLVINCEVDFDQHEHNINGVPPHILGHPHFGG 267
Db 232 EP--VGICQGSTVQSSLRHVPLIQPVKSVTVTCFEFSQAQ-----VP-----G 274
QY 268 HERSSTTKPPFK-----PHGSRDHHHPKHEHGP-----PPPPDERDHDHSHG 309
Db 275 DENPAVTGQPKLPQKNTAPTSPSITAPRGSIQHLPELDDEKPEESKG 323

RESULT 7
KNG_MOUSE
ID KNG_MOUSE STANDARD; PRT; 661 AA.
AC O08677; O08676;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kininogen precursor [Contains: Bradykinin].
GN KNG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS HMW AND LMW).
RC STRAIN=C57BL/6 x CBA; TISSUE=Liver;
RA Takano M., Kondoh J., Yamada K., Okamoto H.;
RT "Molecular cloning of cDNAs for mouse low- and high- molecular
kininogen.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: (1) KININOGENS ARE INHIBITORS OF THIOLESTERASES; (2)
HMW-KININOGEN PLAYS AN IMPORTANT ROLE IN BLOOD COAGULATION BY
HELPING TO POSITION OPTIMALLY PREKALLIKREIN AND FACTOR XI NEXT TO
FACTOR XII; (3) HMW-KININOGEN INHIBITS THE THROMBIN AND PLASMIN-
INDUCED AGGREGATION OF THROMBOCYTES; (4) THE ACTIVE PEPTIDE
BRADYKININ THAT IS RELEASED FROM HMW-KININOGEN SHOWS A VARIETY OF
PHYSIOLOGICAL EFFECTS: (4A) INFLUENCE IN SMOOTH MUSCLE
CONTRACTION, (4B) INDUCTION OF HYPOTENSION, (4C) NUTRIENT AND
DIURESIS, (4D) DECREASE IN BLOOD GLUCOSE LEVEL, (4E) IT IS A
MEDIATOR OF INFLAMMATION AND CAUSES (4E) INCREASE IN VASCULAR
PERMEABILITY, (4F) STIMULATION OF NOCICEPTORS (4E3) RELEASE OF
OTHER MEDIATORS OF INFLAMMATION (E.G. PROSTAGLANDINS), (4F) IT HAS
A CARDIOPROTECTIVE EFFECT (DIRECTLY VIA BRADYKININ ACTION),
INDIRECTLY VIA ENDOTHELIUM-DERIVED RELAXING FACTOR ACTION); (5)
LMW-KININOGEN INHIBITS THE AGGREGATION OF THROMBOCYTES; (6) LMW-
KININOGEN IS IN CONTRAST TO HMW-KININOGEN NOT INVOLVED IN BLOOD
CLOTTING (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; HMW (SHOWN HERE) AND LMW; ARE
PRODUCED BY ALTERNATIVE SPLICING.
```


modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; K02566; AAB35497.1; -
DR EMBL; M11437; AAB59550.1; -
DR EMBL; M11438; AAB59550.1; JOINED.
DR EMBL; M11521; AAB59550.1; JOINED.
DR EMBL; M11522; AAB59550.1; JOINED.
DR EMBL; M11523; AAB59550.1; JOINED.
DR EMBL; M11524; AAB59550.1; JOINED.
DR EMBL; M11525; AAB59550.1; JOINED.
DR EMBL; M11526; AAB59550.1; JOINED.
DR EMBL; M11527; AAB59550.1; JOINED.
DR EMBL; M11528; AAB59550.1; JOINED.
DR EMBL; M11437; AAB59551.1; -
DR EMBL; M11438; AAB59551.1; JOINED.
DR EMBL; M11521; AAB59551.1; JOINED.
DR EMBL; M11522; AAB59551.1; JOINED.
DR EMBL; M11523; AAB59551.1; JOINED.
DR EMBL; M11524; AAB59551.1; JOINED.
DR EMBL; M11525; AAB59551.1; JOINED.
DR EMBL; M11526; AAB59551.1; JOINED.
DR EMBL; M11527; AAB59551.1; JOINED.
DR EMBL; M11528; AAB59551.1; JOINED.
DR PIR; A01279; KGHU1.
DR PIR; A25276; A25276.
DR PIR; A01280; KGHU1.
DR PIR; B25276; B25276.
DR PIR; S02482; S02482.
DR SWISS-2DPAGE; P01042; HUMAN.
DR MIM; 228960; -
DR InterPro; IPR000010; Cystatin.
DR InterPro; IPR003243; Cystatin_C_M.
DR InterPro; IPR002395; Kininogen.
DR Pfam; PF00031; cystatin; 3.
DR PRINTS; PR00334; KININOGEN.
DR Prodom; PD001231; Cystatin_C_M; 1.
DR SMART; SM00043; CY; 3.
DR PROSITE; PS00287; CYSTATIN; 2.
KW Glycoprotein; Plasma; Repeat; Thiol protease inhibitor; Vasodilator;
KW Bradykinin; Blood coagulation; Inflammatory response; Signal;
KW Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 644 KININOGEN.
FT CHAIN 19 380 KININOGEN HEAVY CHAIN.
FT PEPTIDE 381 389 BRADYKININ.
FT CHAIN 390 644 KININOGEN LIGHT CHAIN.
FT DOMAIN 19 136 CYSTATIN-LIKE 1.
FT DOMAIN 137 258 CYSTATIN-LIKE 2.
FT DOMAIN 259 380 CYSTATIN-LIKE 3.
FT DOMAIN 420 510 HIS-RICH (ASSOCIATED WITH CLOTTING ACTIVITY).

FT REPEAT 420 449
FT REPEAT 450 479
FT REPEAT 480 510
FT MOD_RES 19 19 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 28 614 INTERCHAIN.
FT DISULFID 83 94
FT DISULFID 107 126
FT DISULFID 142 145
FT DISULFID 206 218
FT DISULFID 229 248
FT DISULFID 264 267
FT DISULFID 328 340
FT DISULFID 351 370
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 O-LINKED.
FT CARBOHYD 533 533 O-LINKED.
FT CARBOHYD 542 542 O-LINKED.

FT CARBOHYD 546 546 O-LINKED.
FT CARBOHYD 557 557 O-LINKED.
FT CARBOHYD 571 571 O-LINKED.
FT CARBOHYD 577 577 O-LINKED.
FT CARBOHYD 593 593 O-LINKED.
FT CARBOHYD 628 628 O-LINKED.
FT VARSPLIC 402 427 VSPPTSMAPQAQDEERDSKQEGHTR -> SHLRSEYKGR
FT VARSPLIC 428 644 PPKAGAEAPASREVS (IN ISOFORM LMW).
FT CONFLICT 593 593 MISSING (IN ISOFORM LMW).
SQ SEQUENCE 644 AA; 71945 MW; 3132B4CBAF8FB7E CRC64;
T -> I (IN REF. 1).
Query Match 7.1%; Score 217; DB 1; Length 644;
Best Local Similarity 19.4%; Pred. No. 1.8e-07;
Matches 133; Conservative 59; Mismatches 234; Indels 260; Gaps 29;
QY 5 IAALLLITLQYSCAVSPDTCSAVE--PEAKALDLINKRRRGYLFQLLRIADAHLDRIE 62
DB 9 LCSRLLLSLTQESQSEEDCNKDLFKAVDAALKKYNQSNQNNQFVLYRITEA-TKTVG 67
QY 63 NPTVYLVLDVQESDCSVLSRYWNCPEPPDSRR----- 96
DB 68 SDTFYSFKEIKEGDCPVOSGKTWQDCEVKDAAKATGECTATVKRSTKFSVATQTCQ 127
QY 97 --PSE--IVIGQCKVIATRH--SHESQDLRVI-----DFNCTTSVVSALAN----- 137
DB 128 ITPAEGPVVTAQYDCLGCVHPITQSPDLEPLRHGIGYFNNTQHSSLFMLNEVKRAQR 187
QY 138 -----TKDSPVL-----ID-----FFED 150
DB 188 QVAGLNFRITYSVOTNCSENFLFLTPDCKSLWNGDTGECTDAYIDIQRIASFQSN 247
QY 151 TERY-----RKQANKALEKYEKENDDFASPRVDRIERV 183
DB 248 CDIYFGKDFVQPTKICVCCPRDIPITNSPELEETLTHITKLAENNAFFYKIDNVKK- 306
QY 184 ARVGGEGTGYVDFSVR--NCPRHHPRPVNVFGFRA-----DLFYDVEALDLESP 234
DB 307 ARVQVAVGKVFIDFVARETTSKE---SNEELTESCETKILGQSLDCNAEYVVVPEKK 363
QY 235 KNLVINCEVDFQEHENINGVPHLGHFHWG--GHERSTTKPPFKPHGSRDHHHPKP 292
DB 364 IYPTVNCQ---PLGMSLMKRPFGFS-PFRSRIGIEKEETVSP-----P 405
QY 293 HEHGPPPPDERDHSHPPLPQGPPLPMSCSOAHATFTGTAQRHSHNNSSDLHPH 352
DB 406 HTSMAPAQDEERD-----SGKEQGHTR-----HDW 431
QY 353 KHSHEQHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 402
DB 432 GHEKQKKNLGH---GHKHERDQGHGHRGHLGHEQGHGHLGHEQGHGHLGHEQGHG 487
QY 403 GH---HPHCHDFQDYCPDPPPHNOGCHGHPHPPHLLRRRGRGKGPFPFHCRIQSVY 459
DB 488 GHVLDHGH-----KHKHG---HGKHKKNKKNKKNKKNKKNKKNKKNKKNKKNK 531
QY 460 RLPLLRKGE---VLPLPEANPPSPPL-----PHKHPKLPDN----- 493
DB 532 TTPSAQTQETGPTPIPSLAKPGVTVTFSDFQSDSLIATMPPISPAPISQSDDDWIPDI 591
QY 494 -----QPPQSVSESCPGK 507
DB 592 QTDPNGLSFNPISDFPDPTTSPKPCGR 617
RESULT 10
CSUP_DROME
ID CSUP_DROME STANDARD; PRT; 449 AA.
AC Q9V3A4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

FT DOMAIN 136 257 CYPSTATIN-LIKE 2.
FT DOMAIN 258 378 CYPSTATIN-LIKE 3.
FT MOD_RES 19 19 PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .).
FT CARBOHYD 136 136 O-LINKED (PARTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .); OR 169.
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (PARTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .).
FT DISULFID 27 591 INTERCHAIN.
FT DISULFID 82 93
FT DISULFID 106 125
FT DISULFID 141 144
FT DISULFID 205 217
FT DISULFID 228 247
FT DISULFID 263 266
FT DISULFID 327 339
FT DISULFID 350 369
SQ SEQUENCE 621 AA; 68890 MW; D16850BEFE3C55CD CRC64;

Query Match 6.4%; Score 193; DB 1; Length 621;
Best Local Similarity 20.1%; Pred. No. 6.9e-06;
Matches 111; Conservative 61; Mismatches 165; Indels 214; Gaps 29;

Qy 10 LITLOYS-CA-VSP-DCS--AVEPEAEKALDLINKRRRDGYLFOLLRIADAH 57
Db 3 LITLFLCSRLPSLITOESSQEDICNDQDFKAVDAALTKYNSKNGQFVLYRITE-- 60
Qy 58 LDRVEN-TTVYLVLDVQSCSVLSRKYWDCPPDPSRPSEIVIGOCKVIATRHSHES 116
Db 61 VARMDNPDTFSLYQIKEGDCPQSNKTWDCDYKDS---AQATGECTATVAKRGN-- 115
Qy 117 QDLRVDFNCTSSVSSALANTKSPVLIDFFE-----DPE---RYRKQANK 160
Db 116 -----MKFSV--AIQTCLITPAEGPVVTAQYECGLGVHPITKSPDLEPLVRY----- 161
Qy 161 ALEKYEKNDFFASFRVDRIERVARVGEGTGTFVDSV--RNCPRHFF-----PRHPNV 214
Db 162 AIQVFNNTSHSLFDLKEVRAQR-QVYSGWNYEVNYSIAQTNCSEKFEFSFLTPDCKSL 220
Qy 215 ----FGFC-----RADLF-----YDVEALDLSP 234
Db 221 SSGDTGECTAHVDVKLRISFSQKCDLIPVKDFVQPPTRLCAGCPKIPVDSFDLEP 280
Qy 235 -----KMLVINCE 242
Db 281 LSHSLAKLNAEHDGAFYKIDTVKKATVQVAVGLKYSIVFTARETCSKGSNEELTKSC 340
Qy 243 V-----FDQEHENINGVPPHGLHPHGWGHERSSTTKPP----- 277
Db 341 INIHGOILHCDANVYVPEEKVYPTVNCQP--LG-----QTSMLKRPDPGSPFSV 390
Qy 278 --FKPHGSRDHHHPKHEHGGPPPPDSDRSHGPPPLPGQPPPLPMSCSSQHATFCN 335
Db 391 QVMTEGSTTSLPHSAMS---PVQDERDSGK-----EQGP-----THGHWHDH 432
Qy 336 GAQRHSHNNSSDLPHKHEHGHGHPHHAHPHEHDTHRQPHGH--HPHGHHPHGH 394
Db 433 GKQIKLHLGLG---HKH---KHQDGHGHGSHGLGHGKHQKHLGHGKHG--HGK 483
Qy 395 HPHGHHPHGH 405
Db 484 HKNKGKNGKH 494

RESULT 13
KNH2_BOVIN
ID KNH2_BOVIN STANDARD; PRT; 619 AA.
AC P01045;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Kininogen, HMW II precursor (Thiol proteinase inhibitor) [Contains:

DE DE
OS Bos taurus (Bovine).
OC Eukaryota: Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84014106; PubMed=6571699;
RA Kitamura N., Takagaki Y., Furuto S., Tanaka T., Nawa H., Nakanishi S.;
RT "A single gene for bovine high molecular weight and low molecular
weight kininogens.";
RL Nature 305:545-549(1983).
RN [2]
RP SEQUENCE OF 19-376.
RX MEDLINE=87137530; PubMed=3546295;
RA Sueyoshi T., Miyata T., Hashimoto N., Kato H., Hayashida H.,
RA Miyata T., Iwanaga S.;
RT "Bovine high molecular weight kininogen. The amino acid sequence,
positions of carbohydrate chains and disulfide bridges in the heavy
chain portion.";
RL J. Biol. Chem. 262:2768-2779(1987).
RN [3]
RP SEQUENCE OF 376-391.
RX MEDLINE=70180420; PubMed=4986212;
RA Kato H., Nagasawa S., Suzuki T.;
RT "Studies on the structure of bovine kininogen: cleavages of disulfide
bonds and of methionyl bonds in kininogen-II.";
RL J. Biochem. 67:313-323(1970).
RN [4]
RP SEQUENCE OF 387-455.
RX MEDLINE=76260155; PubMed=956151;
RA Han Y.N., Kato H., Iwanaga S., Suzuki T.;
RT "Primary structure of bovine plasma high-molecular-weight kininogen.
The amino acid sequence of a glycopeptide portion (fragment 1)
following the C-terminus of the bradykinin moiety.";
RL J. Biochem. 79:1201-1222(1976).
RN [5]
RP SEQUENCE OF 456-496.
RX MEDLINE=75170265; PubMed=1169237;
RA Han Y.N., Komiya M., Iwanaga S., Suzuki T.;
RT "Studies on the primary structure of bovine high-molecular-weight
kininogen. Amino acid sequence of a fragment ('histidine-rich
peptide') released by plasma kallikrein.";
RL J. Biochem. 77:55-68(1975).
CC -1- FUNCTION: (1) KININOGENS ARE INHIBITORS OF THIOL PROTEASES; (2)
HMW-KININOGEN PLAYS AN IMPORTANT ROLE IN BLOOD COAGULATION BY
HELPING TO POSITION OPTIMALLY PREKALLIKREIN AND FACTOR XI NEXT
TO FACTOR XII; (3) HMW-KININOGEN INHIBITS THE THROMBIN- AND
PLASMIN-INDUCED AGGREGATION OF THROMBOCYTES; (4) THE ACTIVE
PEPTIDE BRADYKININ THAT IS RELEASED FROM HMW-KININOGEN SHOWS
A VARIETY OF PHYSIOLOGICAL EFFECTS: (4A) INFLUENCE IN SMOOTH
MUSCLE CONTRACTION, (4B) INDUCTION OF HYPOTENSION, (4C)
NATRIURESIS AND DIURESIS, (4D) DECREASE IN BLOOD GLUCOSE LEVEL,
(4E) IT IS A MEDIATOR OF INFLAMMATION AND CAUSES (4E1) INCREASE
IN VASCULAR PERMEABILITY, (4E2) STIMULATION OF NOCICEPTORS (4E3)
RELEASE OF OTHER MEDIATORS OF INFLAMMATION (E.G. PROSTAGLANDINS),
(4F) IT HAS A CARDIOPROTECTIVE EFFECT (DIRECTLY VIA BRADYKININ
ACTION, INDIRECTLY VIA ENDOTHELIUM-DERIVED RELAXING FACTOR
ACTION)
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- ALTERNATIVE PRODUCTS: HMW II AND LMW II KININOGEN PRECURSORS ARE
PRODUCED FROM THE SAME GENE AS THE RESULT OF ALTERNATE MRNA
SPLICING. THE SEQUENCES OF BOTH KININOGENS ARE IDENTICAL UP
TO RESIDUE 398.
CC -1- TISSUE SPECIFICITY: PLASMA.
CC -1- PTM: BRADYKININ IS RELEASED FROM KININOGEN BY PLASMA KALLIKREIN.
CC -1- SIMILARITY: CONTAINS 3 CYSTATIN-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC EMBL; V01492; CAA24736.1; -
CC EMBL; V01492; CAA24737.1; ALT_SEQ.
CC PIR; A01282; KGB02.
CC PIR; B29559; B29559.
CC HSP; P01038; I490.
CC InterPro; IPR00010; Cystatin.
CC InterPro; IPR003243; Cystatin_C_M.
CC InterPro; IPR002395; Kininogen.
CC Pfam; PF00031; cystatin; 3.
CC PRINTS; PRO0334; KININOGEN.
CC ProDom; PD001231; Cystatin_C_M; 1.
CC SMART; SM00043; CY; 3.
CC PROSITE; PS00287; CYSTATIN; 2.
CC Glycoprotein; Plasma; Repeat; Vasodilator; Alternative splicing;
KW Thiol protease inhibitor; Bradykinin; Blood coagulation; Signal;
KW Inflammatory response.
FT SIGNAL 1 18
FT CHAIN 19 619 KININOGEN, HMW II.
FT CHAIN 19 376 HEAVY CHAIN.
FT PEPTIDE 378 386 BRADYKININ.
FT CHAIN 387 619 LIGHT CHAIN.
FT DOMAIN 19 135 CYSTATIN-LIKE 1.
FT DOMAIN 136 256 CYSTATIN-LIKE 2.
FT DOMAIN 257 376 CYSTATIN-LIKE 3.
FT MOD_RES 19 19 PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. .).
FT CARBOHYD 136 136 O-LINKED (PARTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .); OR 169.
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (PARTIAL).
FT CARBOHYD 204 204 N-LINKED (GLCNAC. .).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. .).
FT CARBOHYD 400 400 O-LINKED.
FT DISULFID 27 589 INTERCHAIN.
FT DISULFID 82 93
FT DISULFID 106 125
FT DISULFID 141 144
FT DISULFID 205 217
FT DISULFID 228 247
FT DISULFID 261 264
FT DISULFID 325 337
FT DISULFID 348 367
FT VARIANT 398 398 T -> P.
FT VARIANT 401 401 L -> V.
FT VARIANT 454 454 H -> K.
FT SEQUENCE 619 AA; 68710 MW; F04320A8EB0E0DA CRC64;
Query Match 6.3%; Score 190; DB 1; Length 619;
Best Local Similarity 21.2%; Pred. No. 1.1e-05;
Matches 114; Conservative 55; Mismatches 180; Indels 188; Gaps 27;
QY 10 LITLOYSKA-VSPT-----DCS--AVEPEAKALDLINKRRDGYLFOLLRIADAH 57
DB 3 LITLILSRLLPSTOESSQIDNDQDFKAVDAALTKYNSKNQFVLYRITE-- 60
QY 58 LDRVEN-TTVYVLVDVQESCVLSRYWNCDEPPDSRR----- 96
DB 61 VARMDNPOTFYSLKQIKEGDCPFOSNKTWQDCDYKDSQAATGCTATVAKRGNKFSV 120
QY 97 -----PSE--IVIGOCKVIATRH--SHESQDLRLVI-----DFNCTTS 129
DB 121 AIQTCLITPAEGPVVTAQVECLGCVHPITSTKSPDLPLVRLVIAIYFNNTTSHSLFDLKE 180
QY 130 -----SVSSALANTK-----DSPVLI 145
DB 181 VKRAQKQVVGWNYEVNYSIAQTNCSEEFSLTPDCKSLSGDTGECTDKRAHVDVKLRI 240
QY 146 DFFE-----DTERYRKQANKALEYKKEENDDFASFYDR 179
```

```
DB 241 SFSKCDLYPGEDLPPMVCVCCPKPIPVSDPLLEALNHSIAKNAEHDCTFFFKIDT 300
QY 180 IERVAVRGEGTGYFVDFSVR--NCPRIHFFPR-----HPNVFG---FCRADLFYDVEAL 229
DB 301 VKK-ATVQVVGGLKYSIVFIARETTCSKSGNEELTKSCINHGQILHCDANVY---VV 355
QY 230 DLESPKNLVINCEVDFDQEHENINGVPPHLGHPFHGWGHERSSTTKPPKPHGSRDHHHP 289
DB 356 PWEKVIPTVNCQ---PLQOTSIMKRPPGFS-PFR-----SVQVMKTEGSTVSLP 402
QY 290 HKPHEHGGPPPPDERDHSHGPPPLPGPPPLLPWSCSCSOHATFGINGAQRHSHNNSSDL 349
DB 403 HSAMS---PVQDEKDSGK-----EOGP-----THGHWDHGKQIKLHGLGLG-- 442
QY 350 HPHKHSHSEQHPHGHHAHPEHDTTHRQHPHG-HPHGHPHGHGHHPHGHHPHGHGH 405
DB 443 --HKH---KHDQGHGHRSHGLGHGKHQKHGKHGKNGKNGKH 492
RESULT 14
NO75_SOYBN
ID NO75_SOYBN STANDARD; PRT; 309 AA.
AC P08297;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Early nodulin 75 precursor (N-75) (NGM-75).
GN ENOD2A AND ENOD2B.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV WAYNE;
RX MEDLINE=91322483; PubMed=2101308;
RA Franssen H.J., Thompson D.V., Idler K., Kormelink R., van Kammen A.,
RA Bisseling T.;
RT "Nucleotide sequence of two soybean ENOD2 early nodulin genes
RL Plant Mol. Biol. 14:103-106(1990).
RN [2]
RP SEQUENCE OF 69-309 FROM N.A.
RA Franssen H.J., Nap J.-P., Gloudemans T., Stiekema W., van Dam H.,
RA Govers F., Louwerse J., van Kammen A., Bisseling T.;
RT "Characterization of cDNA for nodulin-75 of soybean: a gene product
RT involved in early stages of root nodule development.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4495-4499(1987).
CC -!- FUNCTION: NODULIN N-75 IS INVOLVED IN EARLY STAGES OF ROOT NODULE
CC DEVELOPMENT.
CC -!- INDUCTION: DURING NODULATION IN LEGUME ROOTS AFTER RHIZOBIUM
CC INFECTION.
CC -!- SIMILARITY: TO OTHER PLANTS N-75.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; M16976; AAA33996.1; -
DR EMBL; X16875; CAA34758.1; -
DR EMBL; X16876; CAA34759.1; -
DR PIR; A27059; A27059.
DR PIR; S08343; S08343.
KW Nodulation; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 309 EARLY NODULIN 75.
FT CONFLICT 262 262 Y -> C (IN REF. 2).
FT SEQUENCE 309 AA; 35966 MW; 2773D28CCBDE1654 CRC64;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 31, 2002, 15:02:36 ; Search time 30.96 Seconds
(without alignments)
2933.538 Million cell updates/sec

Title: DAVISN-P04196.PEP
Perfect score: 3036
Sequence: 1 MKALIAALLTLQYSCAVS.....GFKSGFPQVSMFFTHTPPK 525

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvivirus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1901.5	62.6	546	6 Q9BGU1	Q9bgul Bos taurus
2	1750	57.6	525	11 Q99PS6	Q99ps6 mus musculus
3	1746	57.5	525	11 Q9ESB3	Q9esb3 mus musculus
4	1742	57.4	525	11 Q99PS5	Q99ps5 mus musculus
5	1733	57.1	515	11 Q99PS7	Q99ps7 rattus norv
6	1725	56.8	525	11 Q99PS8	Q99ps8 rattus norv
7	1705.5	56.2	510	11 Q9ESB2	Q9esb2 rattus norv
8	667	22.0	214	6 Q9TS85	Q9ts85 bos taurus
9	269	8.9	735	5 Q9NES7	Q9nes7 caenorhabdi
10	266	8.8	140	5 Q26056	Q26056 plasmodium
11	256	8.4	594	5 Q9VEP4	Q9vep4 drosophila
12	238.5	7.9	294	2 Q9K187	Q9ki87 bacillus ce
13	232	7.6	623	5 P90533	P90533 dictyosteli
14	229.5	7.6	296	2 Q9K189	Q9ki89 bacillus ce
15	223.5	7.4	471	5 Q9UAY0	Q9uay0 caenorhabdi
16	222.5	7.3	372	5 Q9VSZ8	Q9vsz8 drosophila

17	217	7.1	485	5	076876	076876 drosophila
18	216.5	7.1	273	2	Q9KI88	Q9ki88 bacillus ce
19	213.5	7.0	102	5	Q9VUE1	Q9vuel drosophila
20	213	7.0	774	3	042918	042918 schizosacch
21	209.5	6.9	1002	3	013849	013849 schizosacch
22	208	6.9	1033	10	Q9XEK5	Q9xek5 chlamydomon
23	203	6.7	403	16	Q98H88	Q98h88 rhizobium l
24	202.5	6.7	940	4	Q96QC0	Q96qc0 homo sapien
25	202.5	6.7	1571	11	054978	054978 mus musculu
26	201.5	6.6	130	2	Q9KI77	Q9ki77 bacillus an
27	201.5	6.6	133	2	Q9KI76	Q9ki76 bacillus an
28	201.5	6.6	265	2	Q9KI91	Q9ki91 bacillus an
29	201	6.6	112	11	Q9GR74	Q9gr74 mus musculu
30	200.5	6.6	356	13	073679	073679 brachydanio
31	198.5	6.5	356	13	Q98UK5	Q98uk5 brachydanio
32	198.5	6.5	796	5	017532	017532 caenorhabdi
33	197	6.5	330	10	041402	041402 sesbania ro
34	197	6.5	341	5	Q9VV44	Q9vv44 drosophila
35	197	6.5	815	5	Q9V9W8	Q9v9w8 drosophila
36	197	6.5	865	5	Q9U7D4	Q9u7d4 neospora ca
37	197	6.5	1480	10	Q9LIE8	Q9lie8 arabidopsis
38	196.5	6.5	780	4	Q96P76	Q96p76 homo sapien
39	195	6.4	554	5	Q9W4C1	Q9w4c1 drosophila
40	194.5	6.4	2715	5	061603	061603 drosophila
41	194	6.4	884	4	000302	000302 homo sapien
42	194	6.4	2703	5	Q9VEG7	Q9veg7 drosophila
43	193.5	6.4	350	11	Q63179	Q63179 rattus norv
44	193	6.4	209	11	Q99KB0	Q99kb0 mus musculu
45	193	6.4	778	5	Q965D9	Q965d9 drosophila

ALIGNMENTS

RESULT 1

ID	Q9BGU1	PRELIMINARY;	PRT;	546 AA.
AC	Q9BGU1;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)			
DE	HISTIDINE-RICH GLYCOPROTEIN.			
GN	BTHRG.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,			
RA	Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;			
RT	"Molecular diversity of mammalian histidine-rich glycoprotein."			
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB055894; BAB33091.1;			
DR	InterPro; IPR000010; Cystatin.			
DR	Pfam; PF000031; cystatin; 1.			
DR	SMART; SM00043; CY; 2.			
SQ	SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;			

Query Match 62.6%; Score 1901.5; DB 6; Length 546;
Best Local Similarity 63.2%; Pred. No. 2.8e-156;
Matches 349; Conservative 45; Mismatches 125; Indels 33; Gaps 5;

Qy	1	MKALIAAL---LLTLQYSCAVSPTDCSVEPEAKALDLINKRRDGYLFQLLRIADAH	57
Db	1	MKVITALLSLLTLTQCSCAVNPTGDAVEPAVALDLINKRWGDFLQLLRVADAH	60
Qy	58	LDRVENTVYVLVDVQESDCVLSRKYNDCPEPPDSRRPSEIVIGQCKVIATRHSQESQ	117
Db	61	LDRVESTAVYVLVDVRESDCPVLRSKRHWDCELNVSRPSETVIGQCKVIATILLAGPE	120

Qy	118	DLRVIDFNCTTSSVSSALANTKOSPVLIDFFEDTERYRKQANKALEKYKEENDDFASFRV	177
Db	121	DLRVNDFNCTTSSVSSALTNTIDSPVLFYFFEDETLYRQAEKALEKYQRENSDFAPFRV	180
Qy	178	DRIEVARVRGGEGTYFVDFSVRNCPRHHPRHPNVEGFCRADLFYDVEALDLSPKNL	237
Db	181	DKVKEMRARGCEGTSYFLDFSVRNCSSHHFPRHSHIFGFCRADLFYDVEASDLETPKDI	240
Qy	238	VINCEVDFQBEHENINGVPHLGHFPHWGHERSSTTKPFKPHGSRDHHHPKPHHEGP	297
Db	241	VTNCEVENLKHEHRSVAVQHLGRPHFSGHEHSPAGRPFPKPSGSKDGHGPHESYNFC	300
Qy	298	PPPPDERDHSHP-----PLPOGPPDLLPMSCSSCOHATFGTNGAQRSHSHNNSSDLHP	351
Db	301	PPPLEHKNSSDSPFQARALPPPPGL-----RCPRPPFGKGHNRPHPHSSDEHNP	354
Qy	352	HKHNSHQHNRHGHNPHRPHAEHDTNRQHRGHNHRGHNRPHGHNRPHGHNRHCH--	409
Db	355	HGHNPHGHNRHGHNHRGHNRGHQHRGHNHRGHNRPHGHNRPHGHNRPHGHNRHHP	414
Qy	410	-----DFQYGCDCPPPHN---QGHCCGHGHPGPPGHLRRRGFGKGRPPFHCOR	453
Db	415	YGHNRPHGHNRHNDYDGHGCDPPPHRQDPQDHRQGRPPPHRSHKERGFGKHFFYWR	474
Qy	454	QIGSVYRLPPLLRKCEVLPLEAFNPPSPRLPHNRHKLPDNOPPQSVSSCCKFKSGRP	513
Db	475	PTGYIHLRPLSKCEVLPLEAFNPPSPSLPNHNPLQPEIQAFPOASRSCFGTFNIKEL	534
Qy	514	QVSMFFTHTPPK	525
Db	535	HISKFFAYTLPK	546

```

RESULT      2
Q99P56
ID          Q99P56
AC          PRELIMINARY;      PRT;      525 AA.
DT          01-JUN-2001 (TREMBlrel. 17, Created)
DT          01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT          01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE          HISTIDINE-RICH GLYCOPROTEIN.
DE          MMHRG.
OS          Mus musculus (Mouse).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX          NCBI_TaxID=10090;
[1]
RN          SEQUENCE FROM N.A.
RP          STRAIN=BALB/C; TISSUE=LIVER;
RC          Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA          Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT          "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL          Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR          EMBL: AB055897; BAB33094.1.
DR          InterPro: IPR000010; Cystatin.
DR          Pfam: PF00031; cystatin; 1.
DR          SMART: SM00043; CY; 2.
SQ          SEQUENCE      525 AA;      59090 MW;      A83E93A439CFB3AC CRC64;
Query Match      57.6%;      Score 1750;      DB 11;      Length 525;
Best Local Similarity      62.3%;      Pred. No. 3.5e-143;
Matches 335;      Conservative      51;      Mismatches 126;      Indels      26;      Gaps      9;

```

Db	119	VNGVNCNTSSVSSALRNTKDSVLDDFFEDSELYRKQARKALDKYKTNDGDFASFRVERA	178
Qy	181	ERVARVRGGEGTGfYDfSVNCRPHHfPVPHPNVFGCRADLfYDVeALDLSPKLNlVIN	240
Db	179	ERVIRARGGETNYIVFESMRNCSTQHfPRSPLVFGCRALLSVfSfTSLDTfPDSfIDIN	238
Qy	241	CEVFDPQEHENINGVPPHLGHfPFHWGGHERSfTTkPfPKfPGHSRDHHPhKHPhHGHPppP	300
Db	239	CEVNIEDHKDTSdkMP-----HW-GHERPLCDKhLCKLSGRDHHTTKTDKLGCPPP	291
Qy	301	PDERDHSHGPPLPQGPPPLPMSCSSCQHATfGTNGAQRHSHNN--NSSDLH----PKHHK	355
Db	292	PEGKDNSDRPLQEGALPQLPPGYp--PHS--GANRTHRPfSYNHSCNEHPChGHRPhGHh	347
Qy	356	SHEOHP-----HGHHPhAHPhPEHDThROHPHGHHPhGHhGHhGHhGHhGHhGHhCHD	410
Db	348	PHSHHPCGHSHGHHPHGHHPhSHSHCHHPGPfGHPHHPHGHHPhGHhGHhGHhGHhGHd	407
Qy	411	FQDYGPCDPPPHNQ---GHCHCHGPPPGHLRRRGPGKGPPRfHCRIgSVYRLPPLRKg	467
Db	408	FLDYGPCDPSNSQELKGOYHRGvCPfPGHSHRKGCGKGLFPFHQQIGYVYRLPPLNiG	467
Qy	468	EVLPLEANFPFSPLPHKKHPLKPDNQPPfQSVSSESCPGKfKSGfPOVSMfFTHTfTPK	525
Db	468	EVLTLPEANFPFSPLPNCSRLOEIOPfPOTASRSCPGKFESGPQISKfFGYTfPPK	525

RESULT 3

Q9ESB3	ID	Q9ESB3	PRELIMINARY;	PRT;	525 AA.
AC	Q9ESB3;				
DC	Q9ESB3;				
DT	01-MAR-2001	(TEMBLrel. 16, Created)			
DT	01-MAR-2001	(TEMBLrel. 16, Last sequence update)			
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)			
DE	HISTIDINE-RICH GLYCOPROTEIN.				
GN	HRG.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxId=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=129;				
RX	MEDLINE=20307726; PubMed=10849117;				
RA	Hulett M.D., Parish C.R.;				
RT	"Murine histidine-rich glycoprotein: cloning, characterization and				
RT	cellular origin.";				
RL	Immunol. Cell Biol. 78:280-287(2000).				
DR	EMBL: AF194028; AAG28416.1; -.				
DR	InterPro: IPR000010; Cystatin.				
DR	Fram; PF00031; cystatin; 1.				
DR	SMART: SM00043; Cy; 2.				
SO	SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;				

	Query Match	57.5%;	Score 1746;	DB 11;	Length 525;
	Best Local Similarity	62.3%;	Pred. No. 7.7e-143;		
	Matches 335;	Conservative 51;	Mismatches 126;	Indels 26;	Gaps 9;
Qy	1	MKALIAALLITLQYSCAVSPTDCSAVEPPEAEKALDILINKRRRDGYLFOLLRIADAHLDR	60		
Dd	1	MKVLTAAALLVTLQCShALSPGNCDAEPLAEKVLIDLINKRRGGYVFELLRVSDAHLDR	60		
Qy	61	VENTVTYYLVLDVOESCYSLSRKYWYNDCCPPDSRRPSEIIVGICKVIATRHSHESQDLR	120		
Dd	61	AGTATVYYLALDVIESDCWLSTKAQDDCLP-SRWQSEIIVGICKVIATRYSNESQDLS	118		
Qy	121	VIDNCTTSSVSسالANKUSPVLLDFPEDTERYRKQANKALEKYKEENDDFASFRVDRI	180		
Dd	119	VNGYNCTTSSVSسالNKUSPVLLDFDESELYRKQARKALKDYKDNGDFSASFRVERA	178		
Qy	181	ERVARVRGGEGTGTYFVDFFSVRNCPRHHPHPRHNPNVFGFCRADLFDYDVALDLESFKNLVIN	240		

```
Db 179 ERVIRARGERTNYVFEFMRNCSTQHPRSPVLFVGFRCALLSYSIETSDLETDSIDIN 238
Qy 241 CEVFPDPOEHENINGVPPHGLHPFHGWHGHERSSTTKPKPKHGSRDHHPHHPKHPHGGPPP 300
Db 239 CEVFNIEDHKDTSMDK-----HW-GHERPLCDKHLCKLSGSRDHHHTHKTDLGCGPPP 291
Qy 301 PDERDHSHPPLPQCPPLPMSSCSQHATFTGNGAQRHSHNN--NSSDLH---PHKHH 355
Db 292 PEGKNSDRPRLOEGALPQLPPGYP--PHS--GANRTHRPYSNHCNEHPCGHPRPHGHH 347
Qy 356 SHEQHP-----HGHHPHAHHPHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHGHHPHCHD 410
Db 348 PSHHHPGHHGHHHPHSHSHSGHHPHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHCHD 407
Qy 411 FQDYGCPDPPHNO---GHCCHGHPGPPGHLRRRPGKPPHFCROIIGSVYRLPPLRKG 467
Db 408 FLDYGCPDPPNSOELKQYHRGYPGPHGHSKRKPGKGLFPFHQIQGYVYRLPPLNIG 467
Qy 468 EVLPLPEANFPFPLPHHKKHLKPDNQPPQSVSESCPGKFKSGFPQVSMFFTHTFPK 525
Db 468 EVLTLPEANFPFSLPNCNRSLOEIQPFQPTASRSCPGKFESEFPQISKFFGYTPPK 525
```

RESULT 4

```
Q99PS5 ID Q99PS5 PRELIMINARY; PRT; 525 AA.
AC Q99PS5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN (UNKNOWN) (PROTEIN FOR MGC:19088).
GN MMHGR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUP-LIVER, NORMAL. 5 MONTH OLD MALE MOUSE.;
RA Strausberg R.;
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB055898; BAB33095.1; -
DR EMBL; BC011168; AAH11168.1; -
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; Cy; 2.
SQ SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;
```

```
Query Match 57.4%; Score 1742; DB 11; Length 525;
Best Local Similarity 62.1%; Pred. No. 1.7e-142;
Matches 334; Conservative 51; Mismatches 127; Indels 26; Gaps 9;

Qy 1 MKALIAALLLITLOYSICAVSTDCSAVEPEAEKALDLINKRRRGYLFQLLRIADHLDR 60
Db 1 MKVLTALLLTLCQSHALSPNCDAEPLAEKVLDLINKRRRGYVFEELLRVSDAHLDR 60
Qy 61 VENTTVYVLVDVQESDCSVLSRYWNCDEPPDSRRRSEIVIGOCKVIAIRHSHESODLR 120
Db 61 AGTATVYVLVDVQESDCSVLSRYWNCDEPPDSRRRSEIVIGOCKVIAIRHSHESODLR 118
Qy 121 VIDFNCTSSVSSALANTKDSPLIDFEDTERYRKANKALEYKKEENDDFASFRVDRI 180
Db 119 VNGVYVLDVIESDCWVLSLTAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLR 178
Qy 181 ERVARVREGGEGTGTFVDFSVNCRPHRHPPHNPVGFRCRADLFDYDVALDESPLNVLIN 240
Db 119 VNGVYVLDVIESDCWVLSLTAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLR 178
Qy 181 ERVARVREGGEGTGTFVDFSVNCRPHRHPPHNPVGFRCRADLFDYDVALDESPLNVLIN 240
Db 119 VNGVYVLDVIESDCWVLSLTAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLR 178
```

```
Db 179 ERVIRARGERTNYVFEFMRNCSTQHPRSPVLFVGFRCALLSYSIETSDLETDSIDIN 238
Qy 241 CEVFPDPOEHENINGVPPHGLHPFHGWHGHERSSTTKPKPKHGSRDHHPHHPKHPHGGPPP 300
Db 239 CEVFNIEDHKDTSMDK-----HW-GHERPLCDKHLCKLSGSRDHHHTHKTDLGCGPPP 291
Qy 301 PDERDHSHPPLPQCPPLPMSSCSQHATFTGNGAQRHSHNN--NSSDLH---PHKHH 355
Db 292 PEGKNSDRPRLOEGALPQLPPGYP--PHS--GANRTHRPYSNHCNEHPCGHPRPHGHH 347
Qy 356 SHEQHP-----HGHHPHAHHPHEHDTHQHPHGHHPHGHHPHGHHPHGHHPHCHD 410
Db 348 PSHHHPGHHGHHHPHSHSHSGHHPHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHCHD 407
Qy 411 FQDYGCPDPPHNO---GHCCHGHPGPPGHLRRRPGKPPHFCROIIGSVYRLPPLRKG 467
Db 408 FLDYGCPDPPNSOELKQYHRGYPGPHGHSKRKPGKGLFPFHQIQGYVYRLPPLNIG 467
Qy 468 EVLPLPEANFPFPLPHHKKHLKPDNQPPQSVSESCPGKFKSGFPQVSMFFTHTFPK 525
Db 468 EVLTLPEANFPFSLPNCNRSLOEIQPFQPTASRSCPGKFESEFPQISKFFGYTPPK 525
```

RESULT 5

```
Q99PS7 ID Q99PS7 PRELIMINARY; PRT; 515 AA.
AC Q99PS7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN 2.
GN RNHRG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUP-LIVER;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055896; BAB33093.1; -
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; Cy; 2.
SQ SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;
```

```
Query Match 57.1%; Score 1733; DB 11; Length 515;
Best Local Similarity 62.2%; Pred. No. 1e-141;
Matches 332; Conservative 48; Mismatches 126; Indels 28; Gaps 7;
```

```
Qy 1 MKALIAALLLITLOYSICAVSTDCSAVEPEAEKALDLINKRRRGYLFQLLRIADHLDR 60
Db 1 MKVLTALLLTLCQSHALSPNCDAEPLAEKVLDLINKRRRGYVFEELLRVSDAHLDR 60
Qy 61 VENTTVYVLVDVQESDCSVLSRYWNCDEPPDSRRRSEIVIGOCKVIAIRHSHESODLR 120
Db 61 VETATVYVLVDVQESDCSVLSRYWNCDEPPDSRRRSEIVIGOCKVIAIRHSHESODLR 118
Qy 121 VIDFNCTSSVSSALANTKDSPLIDFEDTERYRKANKALEYKKEENDDFASFRVDRI 180
Db 119 VNGVYVLDVIESDCWVLSLTAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLR 178
Qy 181 ERVARVREGGEGTGTFVDFSVNCRPHRHPPHNPVGFRCRADLFDYDVALDESPLNVLIN 240
Db 119 VNGVYVLDVIESDCWVLSLTAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLR 178
Qy 181 ERVARVREGGEGTGTFVDFSVNCRPHRHPPHNPVGFRCRADLFDYDVALDESPLNVLIN 240
Db 119 VNGVYVLDVIESDCWVLSLTAQDDCLP--SRWSEIVIGOCKVIAIRHSHESODLR 178
Qy 241 CEVFPDPOEHENINGVPPHGLHPFHGWHGHERSSTTKPKPKHGSRDHHPHHPKHPHGGPPP 300
Db 239 CEVFNIEDHKDTSMDK-----HW-GHERPLCDKHLCKLSGSRDHHHTHKTDLGCGPPP 291
```



```
QY 415 GPCDPPPHQHCCHGHPGPPGHLRRRGGKGRPFHQICQISVYRLPLRKGEVLPLPE 474
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 GPCDPPSSQYH--QCHGPPHGRKRGKGLFPFHQRIGYVYRLPLNVLNVTPE 459

QY 475 ANFPFPLPHKPLKPDNPQPSVSESCPKGKSGPQVSMFETHFPK 525
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 460 ANFPFSLPNCNRPQPEIRFPQASKSCPKGKFGKFPQVSNFETHPPK 510

RESULT 8
Q9TS85 PRELIMINARY; PRT; 214 AA.
AC Q9TS85:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HISTIDINE-RICH GLYCOPROTEIN-FACTOR XIIIA SUBSTRATE (FRAGMENTS).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=94220160; PubMed=7909439;
RA Haikier T., Andersen H., Vestergaard A., Magnusson S.;
RT "Bovine histidine-rich glycoprotein is a substrate for bovine plasma
RT factor XIIia.";
RL Biochem. Biophys. Res. Commun. 200:78-82(1994).
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
FT NON_TER 1
FT NON_CONS 23 24
FT NON_CONS 91 92
FT NON_CONS 120 121
FT NON_CONS 180 181
FT NON_CONS 198 199
FT NON_TER 214
SQ SEQUENCE 214 AA; 23982 MW; B8989492D6097A35 CRC64;

Query Match 22.0%; Score 667; DB 6; Length 214;
Best Local Similarity 52.0%; Pred. No. 6.5e-50;
Matches 142; Conservative 21; Mismatches 48; Indels 62; Gaps 7;

QY 18 AVSPDSCSAVEPEAEKALDLINKRRRDGYLFQLLRIADAHLDLRVENTVYVLVDVQSD 77
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 AVNPTGDAVEPVAVRALDLINK-----VADALDKVESIAVYVLVDVQKESD 48

QY 78 CSVLSRKYNDCPEPPDRPSEIVIGQCKVIATRHSHESQDLRVTDNCTTSSSSALAN 137
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 49 CPVLSRKHWDDCELNVSRYPSETVIGQCKVIATITLAGPEDLR-----TVGEYTEGAN 101

QY 138 TK--DSPVLDFEDTEREYRKQANKALEKYRENDDFASFRVDRTERVARVGGEGTYF 195
|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 102 TRIDSPVLFEDDTLY-----GGEGTSYF 128

QY 196 VDFSVRNCPRIHFRPHPNVFGCRADLFYDVEALDPSKPLVINCEVDFDQEHENINGV 255
:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 LDFSVRNCSHHFPHSHIFGCRADLFYDVEASDLPTPKDVTNCFVN---LKNFSAV 185

QY 256 PPHLGHPHMGCHERSSTTKPPFKPHGSRDHHH 288
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186 QHHLGRPFHWGGH---GNHRPPH---DNSSDEHH 213

RESULT 9
Q9NES7 PRELIMINARY; PRT; 735 AA.
AC Q9NES7:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE Y39B6B.GG PROTEIN (Y39B6A.PP PROTEIN).
GN Y39B6B.GG OR Y39B6A.PP.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Sulston J.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL132896; CAB60938.1;
DR EMBL; AL132948; CAC51077.1;
DR HSSP; P13231; IHCE.
DR InterPro; IPR002395; Kininogen.
DR PRINTS; PR00334; KININOGEN.
SQ SEQUENCE 735 AA; 80255 MW; 6EF7B831DBE41159 CRC64;

Query Match 8.9%; Score 269; DB 5; Length 735;
Best Local Similarity 28.6%; Pred. No. 8.9e-15;
Matches 80; Conservative 12; Mismatches 100; Indels 88; Gaps 16;

QY 248 EHENINGVPHLG----HPFHWGHERSSTTKPPKPHG----SRDHHHPKPHGHPPP 299
||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 408 EHHGTHHSPAHHGEHGTTHGHHGHHHAPAHGHHGHSHGHHGHHGHHGHHGHHAPAH 467

QY 300 PDERDHSHPPLPQGPPLLPMSCSQCHATFTNGAQRHSHNNSSDLHPKHSHSEQ 359
:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 468 HGHGHHHAP-----AHGHHGHEGTHGHHGSHHSPAH-HGHHGHHH 510

QY 360 HPHGHHPH-----AIHHPH--EHDTHRQHPHHPH---HGHHPHGHHP 396
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 511 HAPAHGHHGHEGTHGHHGHHHAPAHGHHGHEGTHGHHGSHHSPAHGHHGHHHHA 570

QY 397 ---HGHHPHG-----HHPCHDFQDYGCDPPPHNQHCCHG-HGPPCHLRRRPGKGP 447
|||||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 571 PAHHGHHGHHGSHGVHHGH-HESHGHHGHHAPAH--GH--HGERGVHHGH---HGAGYGA 622

QY 448 RPFHCRQIGSVYRLPLRKGEVLPLPEANFPSPFLPHHKH 487
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 623 ---HHGHHGAHH-----HAPHHEH 639

RESULT 10
Q26056 PRELIMINARY; PRT; 140 AA.
ID Q26056:
AC Q26056:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HISTIDINE-RICH PROTEIN (FRAGMENT).
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86174893; PubMed=3007981;
RA Irving D.O., Cross G.A.M., Feder R., Wallach M.;
RT "Structure and organization of the histidine-rich protein gene of
RT Plasmodium lophurae.";
RL Mol. Biochem. Parasitol. 18:223-234(1986).
DR EMBL; M15317; AAA29616.1; -
```

[illegible]

[illegible]

RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

[illegible]

Search completed: May 31, 2002, 15:06:13
Job time: 217 sec

```

!!AA_SEQUENCE 1.0
ID   HRG_HUMAN          STANDARD;          PRT;          525 AA.
AC   P04196;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histidine-rich glycoprotein precursor (Histidine-proline rich
DE   glycoprotein) (HPRG).
GN   HRG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
[1]
RN   RP
RP   MEDLINE=86216149; PubMed=3011081;
RA   Koide T., Foster D.C., Yoshitake S., Davie E.W.;
RT   "Amino acid sequence of human histidine-rich glycoprotein derived
RT   from the nucleotide sequence of its cDNA.";
RL   Biochemistry 25:2220-2225(1986).
[2]
RN   RP
RP   SEQUENCE FROM N.A.
RA   Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN   RP
RP   SEQUENCE OF 214-247 FROM N.A.
RA   Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,
RA   Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluff C.;
RT   "Evidence for the absence of intron H of the histidine-rich
RT   glycoprotein (HRG) gene: genetic mapping and in situ localization of
RT   HRG to chromosome 3q28-q29.";
RL   Genomics 19:195-197(1994).
[4]
RN   RP
RP   SEQUENCE OF 19-27.
RC   TISSUE=Plasma;
RX   MEDLINE=93092937; PubMed=1459097;
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992)
CC   -!- FUNCTION: THE PHYSIOLOGICAL FUNCTION IS NOT YET KNOWN. IT BINDS
CC   HEME, DYES AND DIVALENT METAL IONS. IT CAN INHIBIT ROSETTE
CC   FORMATION AND IS KNOWN TO INTERACT WITH HEPARIN, THROMBOSPONDIN,
CC   AND THE LYSINE-BINDING SITE OF PLASMINOGEN. ON THE BASIS OF ITS
CC   HOMOLOGY WITH HMW KININOGEN, THE HIS-RICH REGION OF THIS PROTEIN
CC   MAY MEDIATE THE CONTACT ACTIVATION PHASE OF INTRINSIC BLOOD
CC   COAGULATION CASCADE.
CC   -!- DOMAIN: IN ADDITION TO HAVING A HIGH HIS AND PRO CONTENT, THIS
CC   PROTEIN HAS MANY INTERNAL REPEATS. 12 TANDEM REPEATITIONS OF A 5-
CC   RESIDUE SEQUENCE (GHHPH, CONSENSUS) FORM A HISTIDINE-RICH REGION.
CC   -!- SIMILARITY: CONTAINS 2 CYSTATIN-LIKE DOMAINS.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M13149; AAA52694.1; -.
DR   EMBL; AB005803; BAA21613.1; -.
DR   EMBL; Z17218; CAA78925.1; -.
DR   PIR; A01287; KGHUGH
DR   SWISS-2DPAGE; P04196; HUMAN.
DR   MIM; 142640; -.
DR   InterPro; IPR000010; Cystatin.
DR   Pfam; PF00031; cystatin; 1.
DR   SMART; SM00043; CY; 2.
KW   Glycoprotein; Heparin-binding; Repeat; Signal.
FT   CHAIN              19          525      HISTIDINE-RICH GLYCOPROTEIN.
FT   DOMAIN             19          136      CYSTATIN-LIKE 1.
FT   DOMAIN             137         254      CYSTATIN-LIKE 2.
FT   DOMAIN             276         321      PRO-RICH.
FT   DOMAIN             350         497      PRO/HIS-RICH.
FT   DISULFID           24          504      BY SIMILARITY.
FT   DISULFID           78          89      BY SIMILARITY.
FT   DISULFID           105         126      BY SIMILARITY.
FT   DISULFID           203         417      BY SIMILARITY.
FT   DISULFID           218         241      BY SIMILARITY.
FT   CARBOHYD            63          63      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD            87          87      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD           125         125      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD           344         344      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD           345         345      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE 525 AA; 59578 MW; A2B124D6CE93114F CRC64;
P04196 Length: 525 May 31, 2002 14:35 Type: P Check: 9896 ..
1  MKALIAALLL ITLOYSCAVS PTDSCAVEPE AEKALDLINK RRRDGYLFQL
51  LRIADAHLDL R VENTTVYLV LDVQESDCSV LSRKYWNCDE PPDSSRRPSEI
101 VIGQCKVIAT RHSHESQDLR VDFNCTTSS VSSALANTKD SPVLIDFFED
151 TERYRKQANK ALEKYKEEND DFASFRVDRI ERVARVRGGE GTGYFVDFSV
201 RNCPRHFFPR HPNVFGFCRA DLFYDVEALD LESPKNLVIN CEVFDPOEHE
251 NINGVPPHLG HPFWGGHER SSTTKPPFKP HGSRDHHRPH KPHEHGGPPPP
301 PDERDHSRGP PLPOGPPPLL PMSCSSCQHA TFGTNGAQRH SHNNNSDLH
351 PKHHHSHEQH PHGHHPHAAH PHEHDTTHQH PHGHHPHGHH PHGHHPHGHH
401 PHGHHPHCHD FQYXGCDPP PHNQGHCHG HGPPPHGLLR RGPKGPRPF
451 HCRQTGSYVR LPPLRKGEVL PLPEANPPSF PLPHHKKHLK PDNQFPFQSV
501 SESCPGKFKS GFGQVSMFFT HTPPK

```

THIS PAGE BLANK (USPTO)